

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Evolution of the Environmental Stress Response in Budding Yeast

Permalink

<https://escholarship.org/uc/item/1z19s9pd>

Author

Heineike, Benjamin Murray

Publication Date

2019

Peer reviewed|Thesis/dissertation

Evolution of the Environmental Stress Response in Budding Yeast

by
Benjamin Heineike

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Hana El-Samad

Hana El-Samad

52F8E320792C4C4...

Chair

DocuSigned by:

Patricia Babbitt

Patricia Babbitt

DocuSigned by:

Alexander Johnson

Alexander Johnson

E2BBE6A63C2745B...

Committee Members

Copyright 2019
by
Benjamin Murray Heineke

Acknowledgments

I have been the beneficiary of generous support, both scientific and moral, throughout my time at UCSF. That is not to mention the people who prepared me to be able to come here to become a scientist. There is a long list of educators from elementary and high school, through the US Naval Academy, Cambridge, and Oxford who prepared me intellectually and academically to take on this challenge. Along the way there were also coaches, scout leaders, family friends, senior officers, chiefs, sailors, colleagues and students who passed on lessons about how to work within a team, push myself, give and receive feedback, and to take pride in my work.

I also have been fortunate to have a loving and supportive family. My mom, Marilyn Fisher, and my dad, Ben Heineike, were not well suited for each other, but they were both wonderful parents to me and my brother Dan, who put us and our education first, loved us and loved helping us explore the world.

My dad was a Marine rifle instructor, a craps dealer in a Tahoe casino, a bartender, a carpenter and an adventurer. He was always ready to take a cross country road trip in an old Van to visit friends and family and take some time camping and fishing. He taught us the work ethic, grit, and sense of humor that was passed on to him by the Heineike family from rural Arkansas.

My mom was a maker of walking sticks, a switchboard manager who gave up a promising career to stay home with us when we were young, a blackjack and roulette dealer, and a person who took pride in providing good service to her customers in whatever capacity. She

taught us the warmth, courage, integrity and ambition that was passed on to her by the Fisher family

Both of them got to see me start my studies here at UCSF, but neither of them saw me finish. A combination of bad luck and some life choices that put fun over health cut their lives shorter than I would have liked, but they did get to live full lives, and I am glad I was able to study here in California and be close to them (in the case of my mom, in the same house for a few years) during their final years.

I also have an amazing wife who has been supportive for me during this long journey that must sometimes feel a bit quixotic to someone who has been able to build such tangible and useful software (and software teams). She has borne the burden of making sure that our family can cover the rent in this expensive town while I have taken a bit too long pursuing my curiosity. She listens patiently while I pass on my Naval Leadership advice to her as she shares the stress and frustration of working in a startup. I respect her integrity, courage, and intellect deeply, and am so lucky to be able to share my life with someone so kind, sensible and fun.

Amy has brought so much into my life, and not least her parents, Trevor and Rachel Horton. They are such warm and generous people and a model for a long and loving marriage. They have helped us immensely (even when they have many others back in England who rely on them) as we have made it through this chapter of our life, and I only hope that we can begin to provide at least as much support as we have received from them when we start the next chapter closer to them in the UK. On top of that there is the whole Horton Family which has been a joy to be a part of.

Amy has also brought two lovely boys into my life, Clifford and Lewis. It is so wonderful watching them grow. As I have slowly built up my research, they have learned to crawl, walk, talk, play, and become curious and kind people. It is such a pleasure to take a break each day from the puzzles and deadlines to just play, read, and watch them grow up.

Here at UCSF, the first person I should acknowledge is Hana. She took the chance of

welcoming me into the lab as a student despite some glaring holes in my understanding of biology and some naivete with regards to what it meant to be a scientist. She always encouraged and modeled curiosity, questioning assumptions, and creativity. She set high expectations for me as she does for everyone in the lab, but also provided the guidance and environment for us to meet those expectations.

Something that I appreciate most about Hana, which my time in the Naval Academy and Navy helped me to recognize, is that she is a person of integrity. From putting her mentee's self-interests above her own, to making sure that accurate data is prized over nice looking results, to ensuring that the lab is a responsible steward of its funding, she is a person who can be trusted explicitly and who does not take her word lightly. Another thing that I appreciate about Hana which is informed by my Navy background is her patriotism. She was not born in America, but she recognizes what is best about our country as well as what needs improvement, and does more than her part to give back to the country through science and involvement with important movements. It is impressive to see her commitment to service all while balancing the many other commitments of the lab and her family.

Approximately every two weeks since I joined the lab, I have met with Hana to discuss my progress on my project and to get advice on how to proceed. Hana has an uncanny ability to spot a sloppy or unclear point, and she also has an amazing ability to imagine important new directions of inquiry. I always walk away from those meetings with a clearer purpose, and often with an exciting vision of some future project that I may not ever get to, but which helps make the work more fun.

Those meetings, along with my lab meeting presentations, program talks, and thesis committee meetings have provided the rhythm of my time through graduate school, and UCSF provides an amazing scientific backdrop for asking and answering important biological questions.

In the lab, the senior students and post-docs in the lab were invaluable mentors when I first got started and still, especially David Pincus and Jacob Stewart-Ornstein. I am so

glad that they are both starting their own labs – they are and will continue to be amazing scientists and I am fortunate to have benefitted from their mentorship.

Susan Chen, Patrick Harrigan, and Graham Heimberg all started in the lab at the same time as me and have been great friends through the journey. From pitching our quals proposals, to challenging each other at lab meeting, to developing protocols together, to helping each other get through the valleys and celebrating the summits, it has been much more fun becoming a proper scientist with those three than it otherwise might have been.

The entire lab has been supportive, patient in response to questioning, inspirational and fun and I hope that I can keep following their personal and scientific journeys over the years.

I have benefitted greatly from all the other IPQB students and other students here at UCSF who are incredibly diligent, creative and generous with their time. I am especially grateful to have been in the same cohort as John Haliburton, and incredibly sharp, down to earth and humble scientist, and a good friend. He and I both started graduate school late (although he had a much better idea of what he was getting into and what he wanted out of it) and he and I both started families while we were here. It has been great to have him as a friend to go through it all with.

I have received mentorship from number of others throughout UCSF while here and I should mention them. Wallace Marshall and his lab took me on for the summer before I joined and provided an exciting and creative place to get my first taste of UCSF. Joe DeRisi put a lot of effort into making the IPQB curriculum challenging and exciting and also took me on as a rotation student and followed my progress through the years. Hao Li was on my quals committee and was always great to chat with about science in the hallway or at the various meetings. Hiten Madhani taught part of the first-year curriculum and was always available for fun and enlightening science discussions. Wendell Lim organized the iGEM program and in addition to providing an example of an extremely busy top researcher taking the time to develop the next generation of scientists, gave myself and other PhD students an opportunity to mentor the iGEM students alongside him.

I owe special thanks to Patsy Babbitt who took me on as a rotation student, supported my NDSEG application by allowing me to develop a very cool project proposal with her, and was on my thesis committee, always giving wise, clear headed and supportive advice.

I also owe a special thanks to Sandy Johnson who, along with Dave Morgan, taught the evolution minicourse that got me excited about researching evolution. Sandy agreed to be on my thesis committee and welcomed me as a guest into his lab to learn the careful and exciting methodology they use to pursue evolutionary questions using yeast as a model organism. It has also been valuable to absorb some of his encyclopedic knowledge of evolution, genetics and cell biology as well as science history through his lab's journal clubs and summer book clubs. On my committee, it has been great to have him on my side, helping me to keep focus on a research direction that would make an important contribution to the wider scientific discussion. His lab has also made me feel welcome as a guest, always ready to discuss science, come to my lab meeting, share good papers, and to collaborate on bioinformatics and lab protocols.

I have also received specific help with these projects that I should acknowledge. Nan Hao and Erin O'shea provided PKA(AS) strains in *S. cerevisiae*. Snigdha Poddar and Jamie Cate provided a CRISPR gene editing plasmid and protocol for yeast. Ivan Liachko and Maitreya Dunam proved their Pan-ARS plasmids which we used as a backbone for Crispr/Cas9 gene editing in yeast. In the lab, Jacob Stewart-Ornstein started thinking about comparisons in the PKA pathway and Susan Chen did a rotation in the lab on the problem and I thank them for laying some of the groundwork for the question. A number of rotation students and summer students including Eric Wong, Derrick Lee, Derek Britain, Melanie Sylvis, Gabe Reder and Kyle Fowler assisted in this project even if not all of their efforts made the final draft, and we thank them for their hard work, creativity, and enthusiasm. I thank Andrew Prokop for his work adapting the CRISPR gene editing protocol to yeast as a technician in the lab. David Pincus provided assistance with the RNA sequencing protocol and an early sequencing run, and I was also fortunate to have Kieran Mace, Susan Chen, and

Joao Fonseca as a little RNA sequencing community within the lab. Sandy Johnson and his lab have provided valuable advice and examples for the evolutionary aspects of this problem. Specifically, Trevor Sorrells provided protocols, strains and plasmids from the Johnson lab which allowed me to work with *K. lactis*. I also thank Toni Galbadon for a helpful conversation about the Whole Genome Hybridization and gene conversion. I thank Ken Wolfe for sharing advice about using the YGOB algorithm and other helpful advice during the latter stage of the project.

Finally, I owe thanks to the organizations that funded me through my tenure here – the IPQB training grant, the National Defense Science and Engineering Grant, the Veteran’s Administration (through the Post 911 GI bill), the Discovery Fellowship program and the NIH grant that supported our project in the lab. I hope that the work that we did justifies the bet they took on us, and I also hope that the skills I learned during my time here will provide future returns on their investments as I strive to pursue investigations that fulfill our curiosity about the natural world as well as providing tangible benefits to society in my future career.

Evolution of the Environmental Stress Response in Budding Yeast

Benjamin M Heineike

Abstract

Great strides have been made in understanding the evolution of gene regulation by focusing on the budding yeast. The budding yeast have similar lifestyles across a wide timespan of evolution and, like their cardinal member the model single-celled eukaryote, are relatively easy to culture in the lab. The phylum also has a long and growing list of members whose genomes have been fully sequenced. One important direction of study that took advantage of the status of budding yeast as a model phylum was the comparative study of gene expression data across species in response to a variety of stresses. A key result from those investigations was that there was a major shift in the set of genes activated under a broad set of stress conditions, termed the Environmental Stress Response, that coincided with an ancient interspecies hybridization, known as the yeast Whole Genome Hybridization (WGH), which doubled the number of chromosomes in an ancestor of *S. cerevisiae*. Although that result provided the field with a rich dataset and robust framework for identifying evolutionary shifts in gene programs, as well as some specific intriguing examples of shifts of entire programs of similarly regulated genes, The specific nature of the

genetic rewirings that gave rise to these changes have yet to be illuminated.

We took advantage of unique tools for analyzing the PKA pathway, namely the ability to monitor pathway activity through nuclear localization of the transcription factor Msn2, as well as our ability to construct a mutant strain in which PKA could be inhibited directly to compare signaling and transcriptional responses in this pathway across two species that straddle the phylogenetic divide of the WGH, *S. cerevisiae*, and *K. lactis*.

In 2, by tracking nuclear localization of the stress responsive transcription factor Msn2 (or SC.Msn2) and its ortholog in *K. lactis* (KL.Msn2). We observed that, while *S. cerevisiae* senses a switch from glucose to sorbitol as a stress, *K. lactis* does not. We also saw that Nuclear localization of KL.Msn2 in *S. cerevisiae* was responsive to osmotic shock, but it was not responsive to PKA inhibition. Furthermore, replacing the nuclear localization sequence of KL.Msn2 with that of *S. cerevisiae*, resulted in responsiveness to PKA inhibition. By creating a *K. lactis* strain in which PKA could be directly inhibited we were able to see that PKA inhibition does cause KL.Msn2 nuclear localization in its native environment in a *K. lactis* cell. These results showed that nuclear localization of Msn2 in response to PKA inhibition is preserved in *S. cerevisiae* and *K. lactis*, but that the conditions that cause PKA inhibition between species are not preserved, and the specific link between PKA inhibition and Msn2 nuclear localization is altered between species.

In 3, by comparing the transcriptional profiles of *S. cerevisiae* strains in which PKA was inhibited, we identified substantial set of paralogs from the yeast WGH, termed ohnologs that were differentially expressed under PKA inhibition. We determined using similar data collected in *K. lactis* and with data from stress-based gene expression studies in multiple budding yeast species that a high-basal level with no regulation by PKA is likely to be the ancestral phenotype for most of these differentially expressed ohnologs. We then studied the emergence of their responsiveness to PKA inhibition by analyzing patterns of the appearance of Msn2/4 binding sites in the promoter sequences of a range of recently

sequenced budding yeast species. These analyses revealed that regulation by PKA arose in some ohnolog pairs in one of the two parental lineages prior to the WGH, leading to differential expression afterwards. In other examples, regulation by PKA and differential expression appears to have arisen following the WGH. This result provides a clear illustration of the unique opportunities presented by a WGH event for generating functional divergence by bringing together two parental lineages with separately evolved regulation into one species. Importantly, it poses the idea that functional divergence of two paralogs can be facilitated through such regulatory divergence, which can persist even when functional differences are erased by gene conversion

Table of Contents

1	Introduction	1
2	Comparison of Msn2/4 localization and expression phenotypes between <i>S. cerevisiae</i> and <i>K. lactis</i>	5
2.1	Introduction	5
2.2	Results	6
2.2.1	KL.Msn2 responds to osmotic shock and glucose dropout but not to a switch from glucose to sorbitol.	6
2.2.2	KL.Msn2 responds to osmotic shock in <i>S.cerevisiae</i> cells, but only weakly to replacement of glucose by sorbitol and PKA inhibition. . .	8
2.2.3	KL.Msn2 does respond to PKA inhibition in <i>K. lactis</i> cells. Msn2 from <i>S. cerevisiae</i> does as well.	9
2.3	Discussion	10
2.4	Figures	14

3	Paralogs in the PKA regulon take different evolutionary routes to divergent expression in budding yeast	20
3.1	Summary	20
3.2	Introduction	22
3.3	Results	26
3.3.1	A PKA analog sensitive allele in <i>K. lactis</i> allows for comparing the transcriptional response to PKA inhibition with that of <i>S. cerevisiae</i> .	26
3.3.2	The transcriptional responses to PKA inhibition in <i>S. cerevisiae</i> and <i>K. lactis</i> share many similarities but also exhibit clear functional differences.	27
3.3.3	Many ohnolog pairs are differentially activated by PKA inhibition in <i>S. cerevisiae</i>	29
3.3.4	High basal expression and low LFC under PKA inhibition seems to be the ancestral state for most DE_{PKA} genes.	32
3.3.5	The STRE is enriched in the promoters of the genes induced by PKA inhibition in <i>S. cerevisiae</i> , and <i>K. lactis</i>	35
3.3.6	Analysis of STREs in the promoters of DE_{PKA} orthologs across species revealed enrichment of STREs in the ZT branch.	37
3.3.7	EGO2/4 provides an example where gain of STREs in the ZT branch may have given rise to PKA dependence which was conserved in syntenic orthologs following the WGH.	39

3.3.8	The STRE in the promoter for GND2 may have arisen in the ZT branch (similar to EGO2/4) but its evolutionary history is obscured by gene conversion.	41
3.3.9	GPM2/3 illustrates a case in which STREs might have been added following the WGH in the sensu stricto lineage.	42
3.4	Discussion	43
3.5	Figures	48
3.6	Materials and Methods	58
3.6.1	Plasmids and strain construction	58
3.6.2	Microscopy	62
3.6.3	RNA sequencing	63
3.6.4	GO term enrichment analysis	65
3.6.5	Ohnolog and ortholog assignment	66
3.6.6	Microarray data processing	68
3.6.7	Identification of PKA targets	69
3.6.8	Differential expression definition	69
3.6.9	Promoter extraction, motif enrichment, and clustering	70
3.6.10	Phylogenetic trees	71
3.7	Supplementary Tables	72

3.8	Supplementary Figures	76
	Bibliography	98

List of Figures

2.1	Msn2 response to glucose removal and osmotic stress for <i>S. cerevisiae</i> and <i>K. lactis</i>	14
2.2	Msn2 response to various stresses in <i>S. cerevisiae</i> and <i>K. lactis</i>	15
2.3	Msn2 response to switches in carbon sources in <i>S. cerevisiae</i> and <i>K. lactis</i>	16
2.4	Response of the KL.Msn2 reporter in <i>S. cerevisiae</i> to glucose dropout, osmotic shock, and PKA inhibition	17
2.5	Effect of swapping the Nuclear localization sequence on the KL.Msn2 reporter in <i>S. cerevisiae</i>	18
2.6	Nuclear localization response of KL.Msn2 to PKA inhibition in <i>K. lactis</i> cells	19
3.1	<i>S. cerevisiae</i> and <i>K. lactis</i>	49
3.2	The high-LFC ohnolog from DE_{PKA} tends to have lower average basal expression in <i>S. cerevisiae</i> , while the low-LFC ohnolog and their shared ortholog in <i>K. lactis</i> tend to have higher average basal expression.	50
3.3	For DE_{PKA} genes, low LFC and high basal expression is the ancestral phenotype.	52

3.4	The STRE is enriched in the promoters of DE_{PKA} high LFC paralogs, and in the promoters of genes induced by stress in various species.	54
3.5	STRE conservation in promoters of DE_{PKA} orthologs	55
3.6	Examples of the STRE arising in the promoter of the DE_{PKA} high LFC ohnolog either prior to or following the WGH.	58
3.7	PKA inhibition inhibits growth and causes Msn2 nuclear localization in <i>S. cerevisiae</i> and <i>K. lactis</i>	76
3.8	PKA targets <i>S. cerevisiae</i> and <i>K. lactis</i>	78
3.9	Basal expression (rlog) for DE_{PKA} low and high LFC ohnologs and their shared orthologs in <i>K. lactis</i>	80
3.10	Correlation between PKA inhibition and stress gene expression data	80
3.11	Induction of DE_{PKA} genes during growth to saturation and in response to stresses related to PKA inhibition	81
3.12	Comparison of RNA-seq and microarray basal expression data	83
3.13	Basal expression across DE_{PKA} orthologs for 11 budding yeast species.	85
3.14	Different sets of genes are differentially expressed in different post-WGH species. 86	
3.15	The conclusion that PKA induction is the derived phenotype is independent of the species in which DE_{Stress} genes are defined.	87
3.16	The conclusion that high basal expression is the ancestral phenotype is independent of the species in which the DE_{Stress} genes are defined.	89

3.17	Location distribution of STREs in promoters	90
3.18	Distribution of number and location of TATA boxes in promoters	91
3.19	Joint distribution of STREs and TATA boxes	93
3.20	pvalue for STRE presence in DE_{PKA} orthologs	95
3.21	STRE counts for promoters of DE_{PKA} orthologs	96
3.22	Effect of Msn2 on response to PKA inhibition	97

List of Tables

3.1	GO term enrichment for PKA targets in <i>S. cerevisiae</i> and <i>K. lactis</i>	72
3.2	Enrichment for ohnologs in PKA targets in <i>S. cerevisiae</i> and <i>K. lactis</i>	73
3.3	<i>S. cerevisiae</i> strains	74
3.4	<i>K. lactis</i> strains	74
3.5	Plasmids	75

Chapter 1

Introduction

Saccharomyces cerevisiae (*S. cerevisiae*) has had an outsized contribution to science as the model single celled eukaryote to go along with its outsized contribution to human culture as the key ingredient to beer, bread and wine. Several factors have contributed to the choice of many researchers to spend their life's work uncovering the mysteries of this tiny but powerful organism. Because it is easy to culture and easy to freeze interesting strains indefinitely it allows for reproducible experiments in the lab. Because of its uncanny ability to incorporate DNA into targeted locations via homologous recombination and the ease of mating it is easy to generate useful strains, including the first gene deletion library, the first extensive library of GFP tagged proteins, and many others¹. Finally, its status as a model organism acts as a positive feedback loop, continually increasing its value as a model organism. Because much was known about the organism, it became one of the first organisms to have its genome fully sequenced. Because its genome was sequence it was the subject of much of the early research on gene expression with microarrays. Founded in 1995, the *Saccharomyces* Genome Database (SGD), dedicated to curating the scientific community's collective understanding of yeast, is one of the longest running websites on the world wide web² and remains active and expanding today. With the possible exception

of the model bacterium *E. coli*, *S. cerevisiae* may be the most fully understood organism in the world.

There are other organisms that share many of *S. cerevisiae*'s beneficial characteristics in the lab, including some of its close and distant relatives, and over the years researchers have turned towards the entire clade of budding yeast as a model group of species for studying evolution³. Despite the similarity of the conditions that these organisms can be cultured in in the lab, they cover a wide timespan of evolution. *S. cerevisiae* and *K. lactis*, the organism that our comparisons focus on in this study, share a similar amount of amino acid substitutions per site in their genomes as starfish and humans. *S. cerevisiae* and *Y. lipolytica*, a more distant relative, share a similar amount of amino acid substitutions per site as *C. elegans* and humans. Early bioinformatic comparisons between sequence yeast genomes revealed that a Whole Genome Duplication (WGD) had occurred around 100 million years ago in the budding yeast lineage, which is incidentally an idiosyncrasy which makes *S. cerevisiae* less ideal as a model for genetic studies⁴. The genome duplication, more of a boon than a burden for researches using budding yeast to study evolution. A study comparing the GAL3 and GAL1 duplicates in *S. cerevisiae* to their shared ortholog in *K. lactis* revealed the important evolutionary concept of Escape from Adaptive Conflict in which an ancestral protein which performs two conflicting functions as a galacto-kinase and a co-inducer of the pathway has its functions split between two paralogs following a gene duplication. That study and several others have made use of paralogs which originated in the WGD to derive a fuller understanding of evolutionary processes.

Two such studies from Aviv Regev's lab focused on the evolution of transcriptional stress responses in yeast^{5,6}. In that study 15 budding yeast species were subjected to a variety of stresses including growth to saturation, heat stress, osmotic stress and oxidative stress with H_2O_2 . In *S. cerevisiae*, it has long been known that many cellular stresses trigger a small stress-specific transcriptional response coupled to a larger common Environmental Stress

Response⁷. in Thompson et al.⁵ and Roy et al.⁶, major changes in the activated portion of the Environmental Stress Response were identified which coincided with the Whole Genome Duplication (WGD).

The Protein Kinase A (PKA) pathway is one of the master regulators of the of the Environmental Stress Response⁷⁻⁹. The study of this pathway is facilitated by the fact that two paralogous transcription factors downstream of the pathway (MSn2 and Msn4), which incidentally regulate the activated portion of the response, are regulated by shuttling in and out of the nucleus in response to changes in PKA activity^{7,10-13}. Thus, unlike many other kinase signaling pathways in the cell, it is possible to get a fairly reliable real-time readout in single cells.

In chapter 2 of this study we exploit that readout to identify differences in the pathway between species ranging from the types of perturbations which are sensed as stresses by the PKA pathway between species, to differences in the Nuclear localization sequences (NLS) between KL.Msn2 and SC.Msn2 that influence how PKA inhibition is sensed by those two proteins.

As part of that investigation we developed a protocol for conducting CRISPR/Cas9 gene editing in multiple budding yeast species to create a mutant strain of *K. lactis* in which PKA could be specifically inhibited using the ATP analog 1-NM-PP1 (known as a PKA Analog Sensitive (AS) strain). This technique has been used to great effect in the study of the PKA pathway in *S. cerevisiae*^{9,13-15}. In chapter 3 we compared the transcriptomes of PKA(AS) strains of *K. lactis* and *S. cerevisiae*. These comparisons led us to identify a number of paralogs from the yeast WGH that were specifically upregulated in *S. cerevisiae* in response to PKA inhibition. We further discovered that many of those WGH paralogs (or ohnologs) were differentially induced. By comparing transcriptional profiles previously collected^{5,6} and basal expression microarrays¹⁶ for a subset of PKA related stresses in a range of budding yeast species, we were able to show that responsiveness to PKA and low

basal expression were likely to be the derived state while high basal activity and insensitivity to PKA inhibition were likely to be the ancestral state for any given differentially expressed ohnolog pair in which one member was activated by PKA. A further bioinformatic investigation of Msn2 binding sites in the promoters of a large collection of recently sequenced yeast species¹⁷ revealed to us some of the modes by which responsiveness to PKA via the appearance of Msn2 binding sites can arise in ohnologs. We saw examples in which responsiveness to PKA inhibition appeared to occur prior to the WGH in the *Zygomycetes-Torulospora* (ZT) branch of yeast species after that branch split from the branch that was closer to the other parent species in the WGH, the *Kluyveromyces-Lachancea-Eremothecium* (KLE) branch, and we also saw examples in which responsiveness to PKA inhibition appeared to occur after the WGH. In one example, GND1/2 we see what appears to be the result of a gene conversion which homogenizes the protein sequences of the two paralogs, but which, at least in some clades, left differential regulation that evolved prior to the WGH intact. This illustrates a potentially more general phenomenon by which WGH can introduce novelty in evolution.

Our use of these recently available genome sequences illustrates another advantage that budding yeast have in the study of evolutionary phenomena. Although the genomes for a number of *S. cerevisiae*'s relatives have been around since the early days of genome sequencing¹⁸, with improvements in genome sequencing technology all of the known relatives of budding yeast will soon be sequenced¹⁹. Combining this wealth of bioinformatic resources with our improved abilities to make high throughput genetic changes to yeast using technologies such as our Pan-ARS CRISPR plasmid and other techniques^{20–22}, and advances in gene editing technology with improvements in high throughput phenotyping using readouts ranging from growth to single cell and single colony RNA sequencing in yeast^{23–25} will make the budding yeast subphylum a focus of evolutionary studies for years to come.

Chapter 2

Comparison of Msn2/4 localization and expression phenotypes between *S. cerevisiae* and *K. lactis*

2.1 Introduction

One advantage of focusing on the PKA pathway for studying the evolution of gene regulation is that, unlike most kinase pathways, in yeast there is a reliably, real time proxy for PKA inhibition that can be tracked in single cells in vivo. This is the Msn2 transcription factor. When PKA is inhibited, the paralogous transcription factors Msn2 and Msn4 are dephosphorylated, translocate into the nucleus and activate a majority of the genes that make up the activated portion of the ESR^{7,11,12,26}. Important principles of dynamic pathway activity^{13,27}, noise propagation and bet hedging^{12,24,28}, and pathway crosstalk^{29,30} have been investigated using this system as a model signaling pathway. We were initially curious to see whether we could study those dynamic phenotypes across

species by comparing dynamics of Msn2 response to nuclear localization in response to stresses known to cause PKA inhibition in *S. cerevisiae*. As we pursued this line of inquiry we discovered that there were changes in the pathway at all levels, from the way stress is sensed, to the way that PKA inhibition is transmitted to Msn2 in both species. Although we never did end up comparing subtle dynamic phenotypic differences of Msn2 Nuclear localization between species, we did make a number of intriguing findings that highlight some of the ways the signaling pathways can evolve over the long evolutionary distances covered by the yeast WGH.

2.2 Results

2.2.1 KL.Msn2 responds to osmotic shock and glucose dropout but not to a switch from glucose to sorbitol.

In order to compare the role of Msn2 in the environmental stress response between *S. cerevisiae* and *K. lactis*. We constructed *S. cerevisiae* and *K. lactis* strains containing fluorescent reporters of the Msn2 ortholog in each species (SC.Msn2 and KL.Msn2). In addition to having a C-terminal RFP tag, these reporters contained point mutations (C623S for SC.Msn2 and C523S for KL.Msn2) to a conserved residue in the Zinc Finger binding domain which ablate DNA binding to minimize the effect of the reporters on the transcriptional profile of the cell. In *S. cerevisiae* the SC.Msn2 reporter was overexpressed using an ADH1 promoter, and in *K. lactis*, the KL.Msn2 reporter was overexpressed using the KL.RPS25a promoter.

In response to an osmotic shock of 0.5M sorbitol, the reporters localized to the nucleus with similar dynamics in both species (Fig 2.1 red lines). In response to a glucose dropout in which 2% glucose (i.e. at a concentration of 0.11M) is replaced with water, the reporters

both responded with a long transient burst, although the burst in *K. lactis* was shorter (approximately 50 min compared to approximately 70 minutes for *S. cerevisiae*) (Fig 2.1 light blue lines). In addition to removing the carbon source (glucose) from the media, the glucose dropout perturbation is also a hypo-osmotic shock, so it is difficult to parse how much of the response is due to each aspect of the perturbation. In order to correct for osmotic shock, we also transferred the cells into media containing 0.11M sorbitol instead of 0.11M glucose. This is an effective strategy for *S. cerevisiae* which cannot grow on sorbitol as a sole carbon source. We were, unfortunately, unaware that *K. lactis* can grow on sorbitol (also known as D-glucitol) as a sole carbon source³¹. In *S. cerevisiae* it is clear that the hypo-osmotic aspect of the glucose dropout is not necessary for the full SC.Msn2 nuclear localization response; the response to a switch from glucose (the preferred carbon source) to sorbitol (which cannot be used as a carbon source in that species) is just as severe as in a complete glucose dropout (Fig 2.1 dark blue lines). We verified that there was no osmotic shock in *S. cerevisiae* under a switch to 0.11M sorbitol by monitoring the nuclear localization of Hog1, a kinase that moves into the nucleus in response to osmotic stress (Fig 2.3 C). In *K. lactis* all we can say is that the switch from glucose to sorbitol does not cause KL.Msn2 nuclear localization. We suspect that this is because *K. lactis* does not sense this change through its environmental stress response pathway, although it is still possible that there is another signaling pathway that senses the change in permissible carbon sources.

We also saw a muted response in *K. lactis* to replacement of glucose with 0.11M galactose, and with either 0.67% or 2% ethanol (Figs 2.2 and 2.3). Our results for replacement with Glycerol were surprisingly nonlinear, with no localization of KL.Msn2 in response to 1.8% glycerol but a response in both 0.8% and 3% glycerol. This counter intuitive result should be repeated and verified before speculating much further, but it could be the result of osmotic shock for 3% glycerol and a combination of hypoosmotic shock and carbon starvation for 0.8% glycerol.

In *K. lactis*, there was also a muted nuclear localization response to oxidative stress with 1.0mM H_2O_2 (Fig 2.2. Exponentially growing *K. lactis* cells typically metabolize carbon using aerobic respiration which generates reactive oxygen species. This reduced response to H_2O_2 may be the result of *K. lactis* cells having their oxidative stress response machinery active to deal with the already high endogenous levels of reactive oxygen species.

2.2.2 KL.Msn2 responds to osmotic shock in *S.cerevisiae* cells, but only weakly to replacement of glucose by sorbitol and PKA inhibition.

To see how well the activated environmental stress response pathway was preserved between *S. cerevisiae* and *K. lactis*, we tested nuclear localization of KL.Msn2 in the context of *S. cerevisiae* cells. Interestingly we saw that KL.Msn2 localized to the nucleus in with similar dynamics to SC.Msn2 in response to osmotic shock with 0.5M sorbitol, however, the response of KL.Msn2 to replacement of glucose with sorbitol was muted compared to that of SC.Msn2.

This result was puzzling because SC.Msn2 is thought to be a reliable reporter for PKA inhibition, and both a drop in glucose levels and osmotic shock cause dephosphorylation of several putative PKA phosphosites on the SC.Msn2 nuclear localization and export sequences (NLS and NES respectively)¹³. If KL.Msn2 was responsive to PKA inhibition in *S. cerevisiae*, then either the switch from glucose to sorbitol was not causing PKA inhibition or the PKA signal was being counteracted for KL.Msn2 by some other signal emanating from the media switch. On the other hand, if KL.Msn2 was not responsive to PKA inhibition in *S. cerevisiae* then the signal for osmotic stress must be transmitted to KL.Msn2 via some other signaling pathway. To distinguish between these two possibilities, we tested the nuclear localization response of KL.Msn2 to direct inhibition of PKA in an *S.*

cerevisiae strain.

In order to inhibit PKA in *S. cerevisiae* we used a PKA Analog Sensitive (AS) strain in which the PKA catalytic subunits carried point mutations rendering them sensitive to the ATP analog 1-NM-PP1¹³. We saw that KL.Msn2 was not responsive to PKA inhibition in *S. cerevisiae*, implying that the switch from glucose to sorbitol must be going through the PKA pathway, and also that KL.Msn2 must be responding to some other signal besides PKA inhibition in response to osmotic shock (e.g. the Hog1 pathway) 2.4.

It is known that PKA phosphorylates serine residues in the nuclear localization sequence (NLS) of SC.Msn2. We wondered if the inability of KL.Msn2 to translocate to the nucleus in response to PKA inhibition in *S. cerevisiae* cells was a result of changes in those sequences. We therefore constructed chimeric reporter proteins in which the KL.Msn2 NLS was replaced with the NLS from SC.Msn2. We also constructed a chimeric SC.Msn2 reporter in which its NLS was replaced with the NLS from KL.Msn2. We saw that, in response to 4 μ M 1-NM-PP1, the NLS from *S. cerevisiae* was sufficient to enable nuclear localization for KL.Msn2, although at a lower level than is seen for SC.Msn2. We also saw that the NLS from KL.Msn2 caused a strong but not complete reduction in the response of the endogenous Msn2 reporter to PKA inhibition. These results imply that changes in the NLS between *S. cerevisiae* and *K. lactis* are in large part responsible for the inability of KL.Msn2 to translocate into the nucleus in response to PKA inhibition

2.2.3 KL.Msn2 does respond to PKA inhibition in *K. lactis* cells. Msn2 from *S. cerevisiae* does as well.

The fact that KL.Msn2 was unable to localize into the nucleus in response to PKA inhibition in *S. cerevisiae* made us wonder if KL.Msn2 was responsive to PKA in *K. lactis*. In order to test this, we constructed a PKA analog sensitive strain in *K. lactis* using a

CRISPR based strategy (see 3.6.1) and monitored translocation of a KL.Msn2 reporter and an SC.Msn2 reporter into the nucleus.

We saw that KL.Msn2 did respond strongly to PKA inhibition in *K. lactis* cells (Fig 2.6A), in line with other results suggesting that KL.Msn2 was downstream of PKA inhibition³².

Thus the differences in the NLS between SC.Msn2 and KL.Msn2 that prevented KL.Msn2 from responding to PKA inhibition in *S. cerevisiae* were insufficient to prevent the transmission of the PKA inhibition signal to KL.Msn2 in its endogenous context in *K. lactis*.

We were also surprised to find that SC.Msn2 responded to PKA inhibition in *K. lactis* cells in contrast to the lack of response of KL.Msn2 to PKA inhibition in *S. cerevisiae* cells (Fig 2.6B). This implies that we are not seeing a simple case of coevolution in which PKA and Msn2 in both species make coordinated changes so that they are incompatible with their orthologs from the other species, but rather a more asymmetric coevolution in which there are changes in the way that PKA interacts with Msn2 between species, but SC.Msn2, due in part to some property of its NLS, is somehow more sensitive to PKA signals from both species.

2.3 Discussion

Because of the wide evolutionary distance between *S. cerevisiae* and *K. lactis* one must be careful in interpreting differences between the species when the measurements are taken from pathways involving many components. Our observation that KL.Msn2 did not localize to the nucleus in *K. Lactis* cells under a change from glucose to sorbitol while SC.Man2 did in *S. cerervisiae* cells did not, as we originally suspected, turn out to be due to the absence of an interaction between PKA and Msn2 in *K. lactis*. It was in fact a

difference between the way that the two strains sensed and utilized sorbitol.

In pursuing this question we also discovered that KL.Msn2 did not localize to the nucleus in response to PKA inhibition in *S. cerevisiae* cells. Again this led us to suspect that the same would hold true in *K. lactis* cells, especially because there was more conservation between PKA active and catalytic subunits (between 71-79% sequence identity) and the transcription factors (Under 18% sequence identity between SC.Msn2 and KL.Msn2) so that any changes in the interaction we saw would be primarily due to a change in the transcription factor but not in the kinase.

There was a change in the transcription factors, and we were able to identify the NLS sequence as an important part of that change since a KL.Msn2 reporter with just an NLS from *S. cerevisiae* could translocate into the nucleus in response to PKA inhibition in *S. cerevisiae*. There were also changes to the PKA pathway which may be due to a number of factors including sequence changes in the catalytic subunits, changes in their regulation, the fact that there is one more isoform of PKA in *S. cerevisiae*, or even to changes in the other components that interact with PKA as it transmits its signal to Msn2.

In the end the best way to tell whether KL.Msn2 translocated into the nucleus in response to PKA inhibition in *K. lactis* was to inhibit PKA directly in *K. lactis* and measure KL.Msn2 nuclear localization. Once we had the ability to inhibit PKA in *K. lactis* directly we shifted our focus to comparing the global gene expression response to PKA inhibition between the two species (3), however our investigations into the differences in nuclear localization of Msn2 between *S. cerevisiae* left a number of interesting questions unanswered.

First, the dynamics of the response to PKA inhibition seemed to be different between the two species. In *S. cerevisiae* SC.Msn2 stayed in the nucleus as long for the duration of the experiment while in *K. lactis* KL.Msn2 began to return to the cytoplasm before 30 minutes

had elapsed. Further work would need to be done to determine whether this is due to a change in the way that each species imports, exports, or metabolizes the drug, changes in the way that the drug interacts with PKA, or some process that allows KL.Msn2 to overcome complete PKA inhibition and return to the cytoplasm.

Secondly, it would be important to discover what is different in terms of the PKA pathway between *S. cerevisiae* and *K. lactis* that prevents KL.Msn2 from responding to PKA inhibition in *S. cerevisiae* while it responds perfectly well in *K. lactis*. This is made further interesting by the fact that SC.Msn2 responds well to PKA inhibition in both species. Is the strength of PKA activity in some sense lower in *K. lactis* than in *S. cerevisiae* and thus KL.Msn2, which we presume translocates to the nucleus when it is no longer phosphorylated enough by PKA, is only able to translocate to the nucleus in its endogenous environment where PKA activity would, by this hypothesis, be lower. Under that hypothesis SC.Msn2 could potentially be able to translocate into the nucleus for both PKA in *S. cerevisiae* which, by the hypothesis would have generally higher levels of PKA activity, and in *K. lactis* which would have generally lower levels of PKA activity.

Next, just because KL.Msn2 responds to osmotic shock *S. cerevisiae* when it doesn't respond to PKA inhibition in *S. cerevisiae* does not imply that PKA inhibition is not a part of the response to osmotic shock in *S. cerevisiae*. It could be that osmotic shock causes PKA inhibition in *S. cerevisiae*, but that that signal is not passed on to KL.Msn2 in that species even though it would be passed on to SC.Msn2. It is also not necessary that the pathway that triggers KL.Msn2 nuclear localization during osmotic shock in *S. cerevisiae*, would also trigger nuclear localization for SC.Msn2. It would be interesting to work out the details of this alternate osmotic stress signal that gets transmitted to KL.Msn2 under osmotic shock with sorbitol. It also would be interesting to see whether removing glucose without altering the media's osmolarity (say by swapping glucose to a compound that cannot be metabolized by either species such as xylose), would trigger

something akin the full response to a swap from glucose to water.

To investigate our observation that adding an NLS region from SC.Msn2 onto the KL.Msn2 reporter was sufficient to allow a nuclear localization response in response to PKA inhibition, it would be interesting to understand how widespread this phenomenon is in the other species descended from the last common ancestor between *S. cerevisiae* and *K. lactis*. To that end one would need to build Msn2 nuclear localization reporters in *S. cerevisiae* for a number of different species and potentially try to rescue nuclear localization ability with the NLS from SC.Msn2. This investigation combined with a bioinformatic comparison of the Msn2 protein between species could identify the specific differences between the SC.Msn2 NLS and the KL.Msn2 NLS that are important for altering the response to pKA inhibition in *S. cerevisiae*. Along these lines it would also be interesting to perform a similar set of experiments as we did for the NLS chimeras by making chimeric proteins in which that part of Msn2 is swapped between species.

It is difficult to isolate a single evolutionary change using only information from across long pathways in evolutionary distant species because there will inevitably multiple changes that can coevolve making more and more distant proteins incompatible with the environment of its distant cousin. If one is to get at specific evolutionary changes in signaling pathways it will be important to look at the phenomena in a fairly narrow range of species, or as we chose to do, limit the length of the pathway under investigation by either artificially perturbing it, or making measurement closer to the perturbation.

2.4 Figures

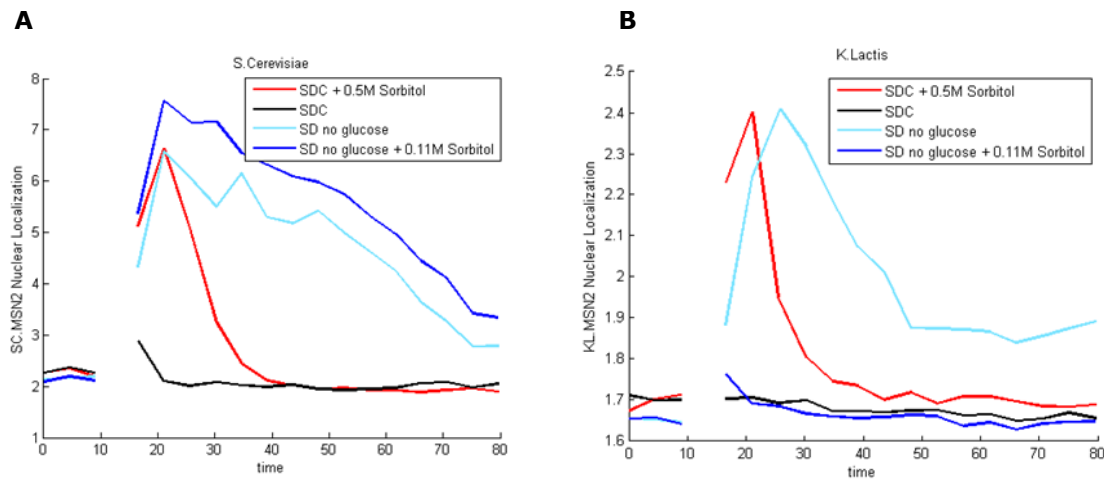


Figure 2.1: **Msn2 response to glucose removal and osmotic stress for *S. cerevisiae* and *K. lactis*** *K. lactis* Msn2 responds to osmotic shock and a complete loss of glucose, but not to a switch from glucose to sorbitol. A) Mean nuclear localization values for single *S. cerevisiae* cells in response to either addition of 0.5M Sorbitol (red), SDC with no glucose (light blue), SDC with no glucose plus 0.11M Sorbitol (equivalent molarity to 2% glucose)(dark blue) or to SDC with 2% glucose (control - black). The gap in each line is a break during which the plate was removed from the microscope and the perturbation was delivered. There was no cell tracking during that gap.

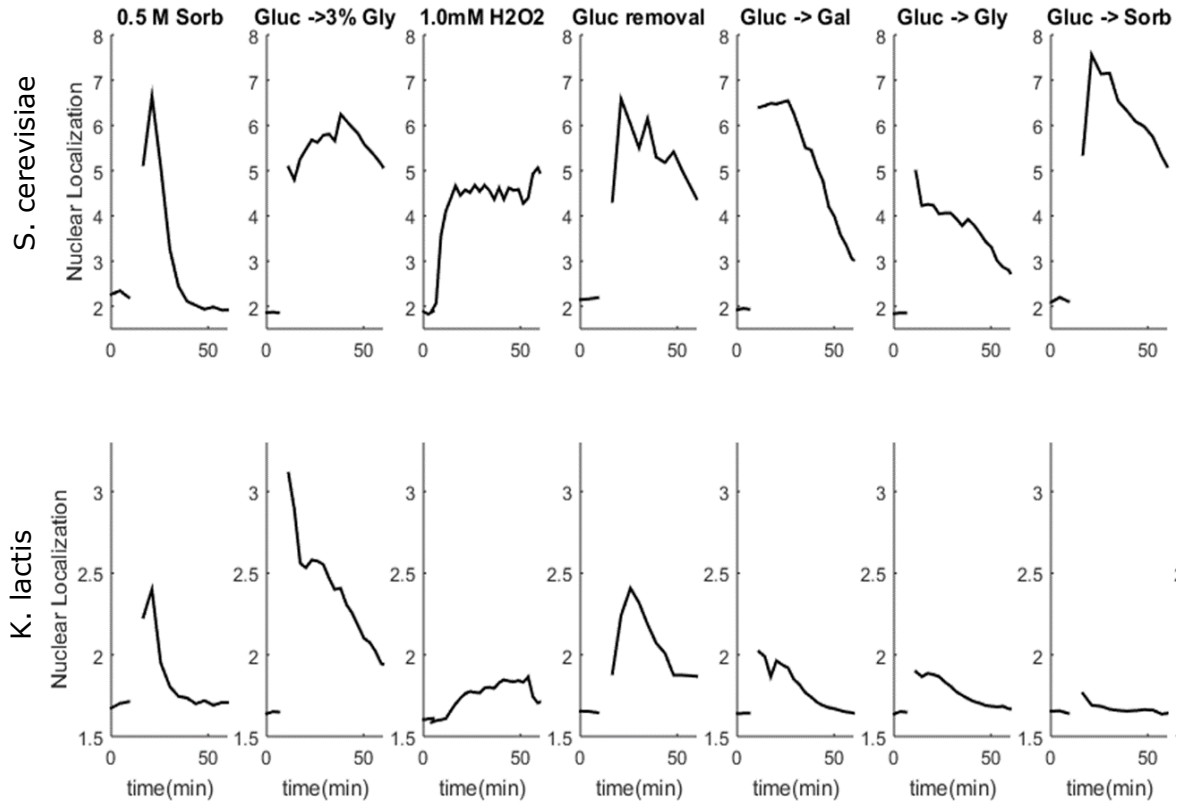


Figure 2.2: **Msn2 response to various stresses in *S. cerevisiae* and *K. lactis*** KL.Msn2 in *K. lactis* responds to osmoshock and glucose removal, but not strongly to H_2O_2 or shifts to galactose, glycerol, or sorbitol as Msn2 in *S. cerevisiae*. Mean nuclear localization of Msn2 in *S. cerevisiae* (top) cells or KL.Msn2 in *K. lactis* (bottom) cells in response to various stress conditions. The gap in each line is a break during which the plate was removed from the microscope and the perturbation was delivered. There was no cell tracking during that gap.

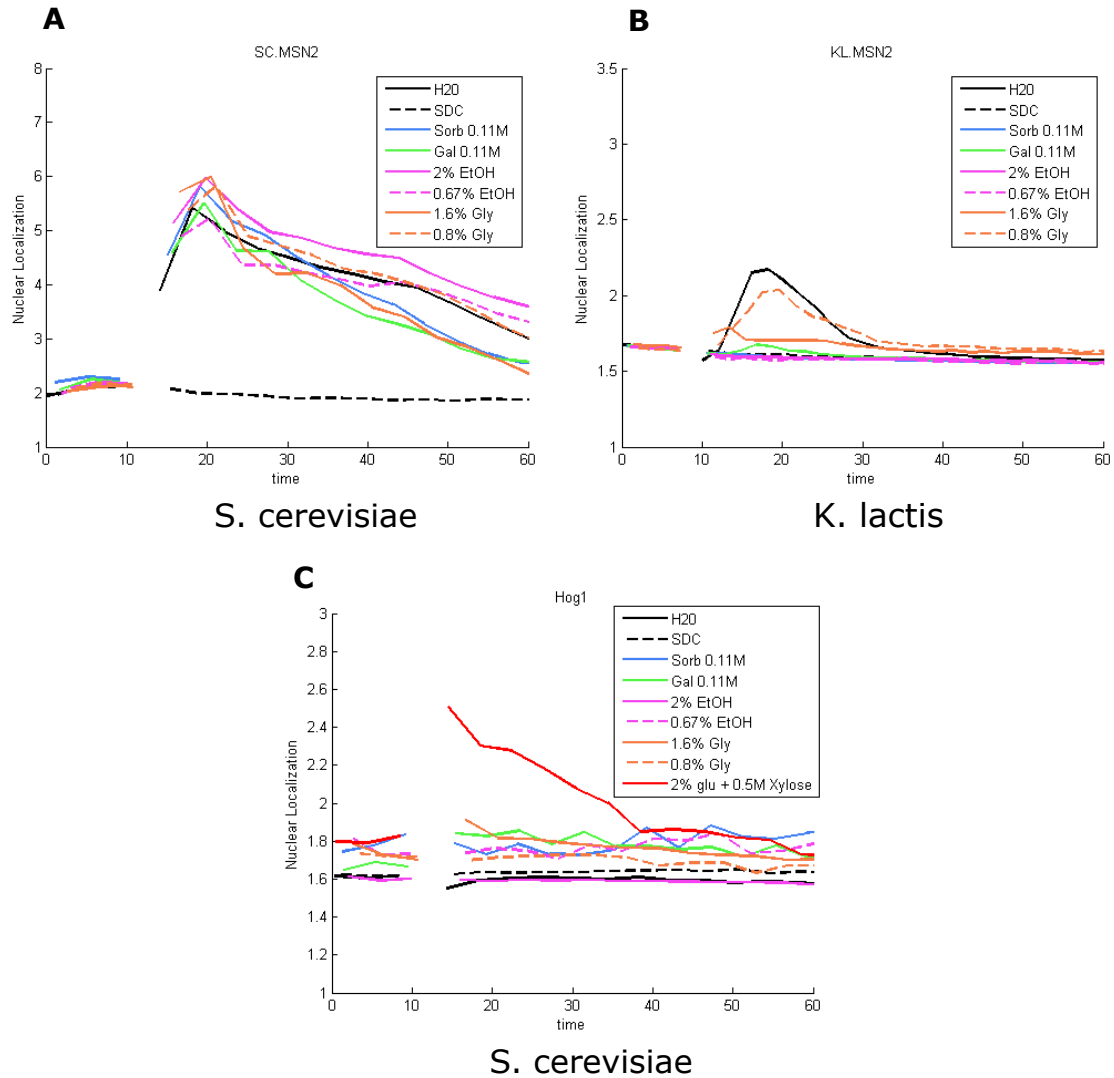


Figure 2.3: **Msn2 response to switches in carbon sources in *S. cerevisiae* and *K. lactis*** KL.Msn2 in *K. lactis* cells is less likely than Msn2 in *S. cerevisiae* cells to respond to a switch in carbon sources. (A) Nuclear localization of Msn2 in *S. cerevisiae* cells in response to a change in carbon sources or to a change to no carbon source (labeled H_2O). (B) Nuclear localization of KL.Msn2 in *K. lactis* cells in response to the same conditions. (C) Nuclear localization of Hog1 in *S. cerevisiae* cells in response to the same conditions in addition to an osmo shock with 0.5M xylose.

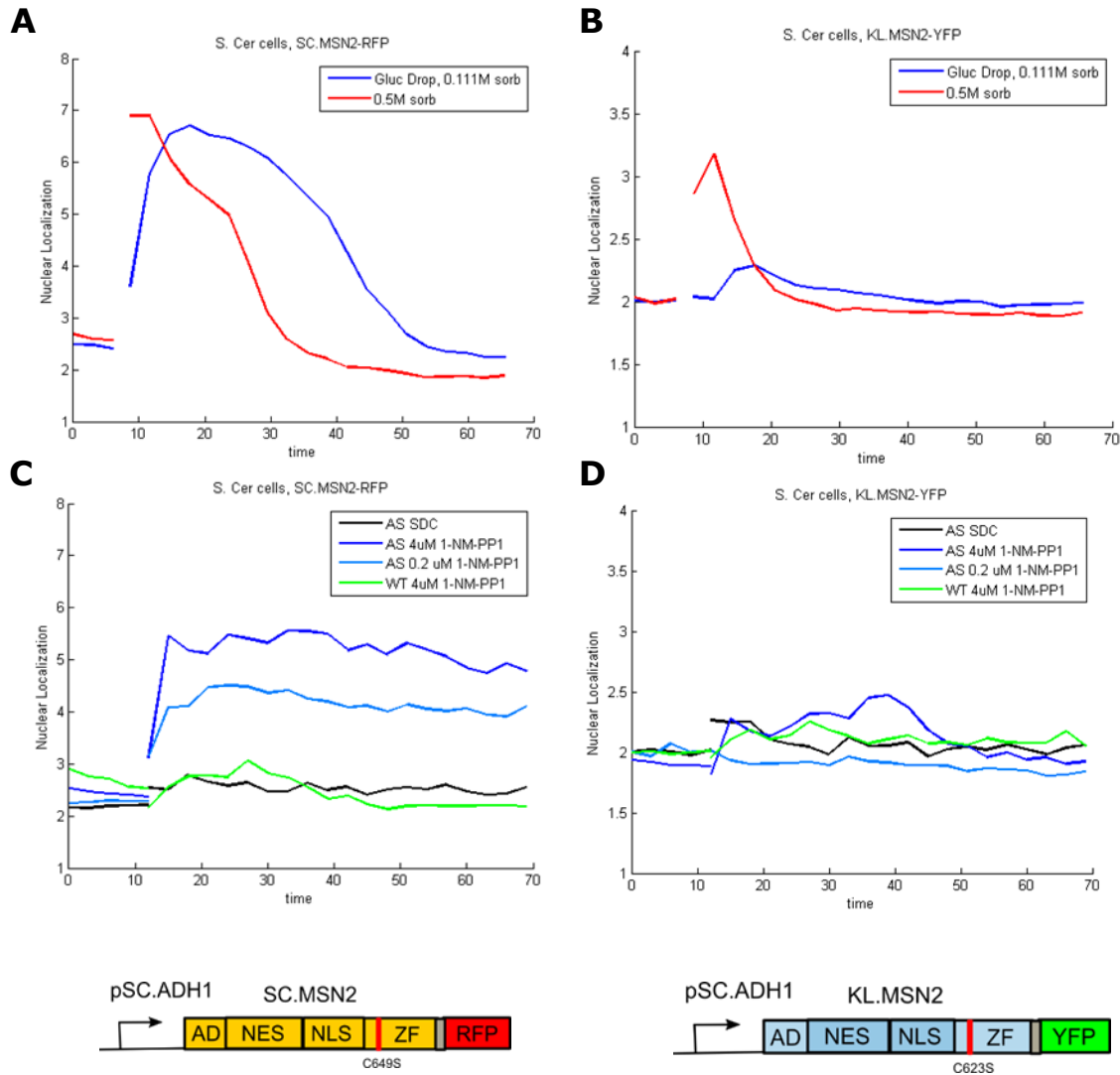


Figure 2.4: Response of the KL.Msn2 reporter in *S. cerevisiae* to glucose dropout, osmo shock, and PKA inhibition KL. Msn2 does not respond to a change from glucose to sorbitol, or PKA inhibition in *S. cerevisiae* cells, but it does respond to osmotic shock. (A) Mean nuclear localization response of Msn2 in *S. cerevisiae* cells to a change from glucose to sorbitol, or to osmotic shock with 0.5M sorbitol. (B) Mean nuclear localization response of KL.Msn2 in *S. cerevisiae* cells to a change from glucose to sorbitol, or to osmotic shock with 0.5M sorbitol. (C) Mean nuclear localization response of Msn2 in *S. cerevisiae* cells carrying a PKA(AS) mutation in response to the ATP analog 1-NM-PP1. The green line represents the response in WT cells to 4 μ M 1-NM-PP1. (D) Mean nuclear localization response of KL.Msn2 in *S. cerevisiae* cells carrying a PKA(AS) mutation in response to the ATP analog 1-NM-PP1. The green line represents the response in WT cells to 4 μ M 1-NM-PP1. The schematics below the figures represent the reporter used in that column for that experiment. The red line in the schematic represents a mutation to the DNA binding domain that removes DNA binding and transcription.

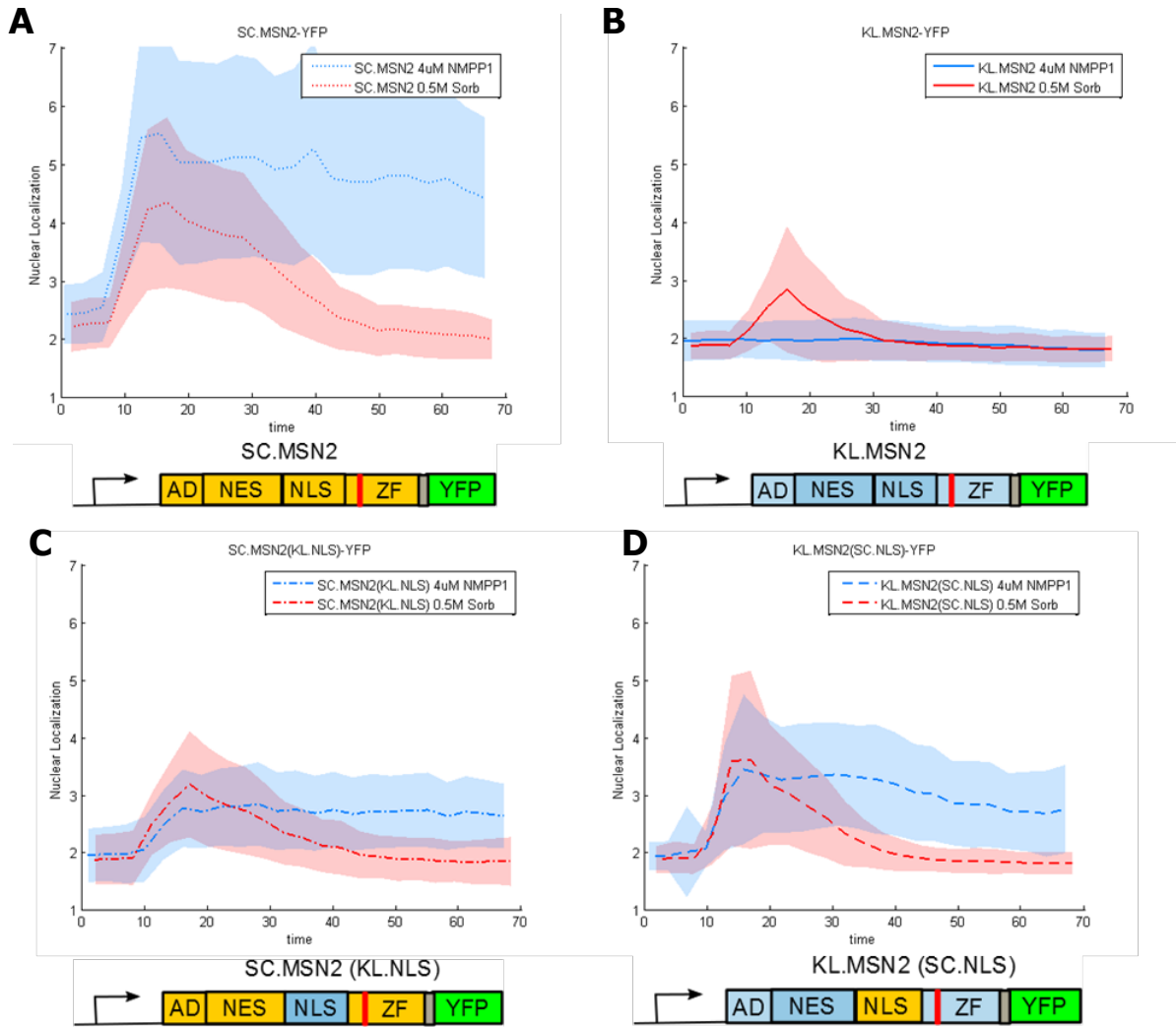


Figure 2.5: Effect of swapping the Nuclear localization sequence on the KL.Msn2 reporter in *S. cerevisiae* Replacing the KL.Msn2 nuclear localization sequence in with that of Msn2 from *S. cerevisiae* renders it able to respond to PKA inhibition in *S. cerevisiae* cells. Mean nuclear localization response in *S. cerevisiae* cells to osmotic shock with 0.5M sorbitol, or to PKA inhibition with 4 μ M 1-NM-PP1. Each subfigure shows results from a different reporter which is described in the schematic below the subfigure as follows (A) Msn2 from *S. cerevisiae*. (B) KL.Msn2 (C) Msn2 (from *S. cerevisiae*) with its nuclear localization sequenced replace by that of KL.Msn2. (D) KL.Msn2 with its nuclear localization sequence replaced by Msn2 (from *S. cerevisiae*). The solid lines represent the mean nuclear localization, and the shaded lines represent the standard deviation for all tracked cells for that datapoint.

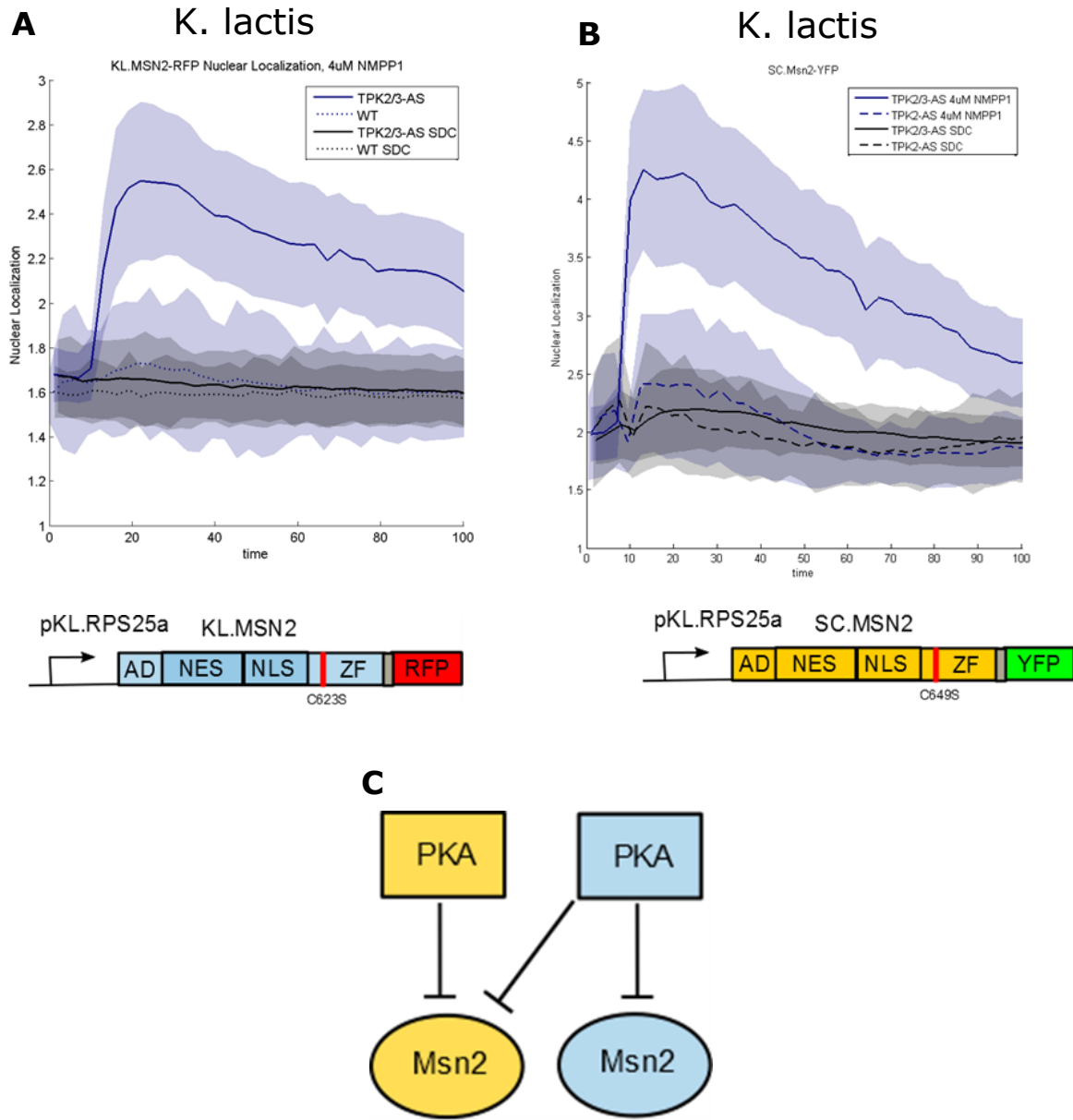


Figure 2.6: **Nuclear localization response of KL.Msn2 to PKA inhibition in *K. lactis* cells** KL.Msn2 responds to PKA inhibition in *K. lactis* cells. Nuclear localization of (A) KL.Msn2 or (B) Msn2 (from *S. cerevisiae*) in response to either control media (black), or 4 μ M 1-NM-PP1 in either WT cells (dashed line) or cells containing a PKA(AS) mutation. Diagrams for reporters used are below each plot. (C) Diagram of the connections between PKA and Msn2 in *S. cerevisiae* and *K. lactis* yellow indicates components from *S. cerevisiae* and blue represents components from *K. lactis*.

Chapter 3

Paralogs in the PKA regulon take different evolutionary routes to divergent expression in budding yeast

3.1 Summary

Functional divergence of duplicate genes, or paralogs, is an important driver of novelty in evolution. In the model yeast *S. cerevisiae*, there are 547 paralog gene pairs that survive from the yeast allo-polyploidization or Whole Genome Hybridization (WGH) that is estimated to have occurred approximately 100 million years ago. Many WGH paralogs (or ohnologs) are known to have differential expression during the yeast environmental stress response, which involves Protein Kinase A (PKA) as one of its major regulators. While investigating the transcriptional response to PKA inhibition in *S. cerevisiae*, we discovered that, of the 148 ohnolog pairs that had at least one member induced by PKA inhibition, 91 had differential expression, comprising approximately 1/6th of all ohnolog pairs. In this

group of ohnolog pairs, one member tended to have low basal expression that increased upon PKA inhibition, while the other member tended to have high basal expression that remained unchanged. The orthologs of these genes in pre-WGH species *K. lactis* tended to have high basal expression, and were not induced under PKA inhibition. Using publicly available expression data for various yeast species under PKA related stress conditions, we determined that a high-basal level with no regulation by PKA is likely to be the ancestral phenotype for most of these differentially expressed ohnologs. To understand how the PKA-regulation phenotype arose, we analyzed gene promoters of these ohnologs for a trademark of PKA regulation — the presence of a binding motif for the PKA responsive transcription factor Msn2 — across a range of sequenced budding yeast species. These analyses revealed different possible evolutionary routes. In some examples, regulation by PKA appears to have arisen in one of the two parental lineages prior to the WGH, leading to differential expression afterwards. In other examples, regulation by PKA and differential expression appears to have arisen following the WGH. Although it was known that many ohnologs were differentially expressed following stress in *S. cerevisiae*, our data are the first to show that much of this differential expression is the result of regulation gained specifically downstream of the PKA pathway following the split between the two parental lineages that contributed to the WGH. More broadly, our studies provide a clear illustration of the unique opportunities presented by a WGH event for generating functional divergence by bringing together two parental lineages with separately evolved regulation into one species. Importantly, it poses the idea that functional divergence of two paralogs can be facilitated through such regulatory divergence, which can persist even when functional differences are erased by gene conversion.

3.2 Introduction

Gene duplication is considered an important source of novelty and adaptation in evolution^{33–37}. In particular, the duplication of entire genomes or chromosomes in polyploidy events, or Whole Genome Duplications (WGD), have been hypothesized to offer a unique avenue to generate evolutionary diversity by allowing certain classes of genes that rarely survive as small scale duplicates to be retained as duplicates after a WGD^{38–41}.

Much of the understanding we have of WGD comes from work done on the model organism *Saccharomyces cerevisiae*, the first eukaryote to have its genome fully sequenced⁴². Not long after the *S. cerevisiae* genome was published, researchers discovered that its lineage had undergone a WGD approximately 100Mya⁴. Today, *S. cerevisiae* retains 547 paralogs from this event, meaning that about 1/6 of all annotated protein coding genes are paralogs from the WGD, termed “ohnologs”¹⁸. As more budding yeast sequences have become available in recent years, greater understanding has developed surrounding the budding yeast WGD. One particularly surprising recent discovery was that the WGD event was most likely an allo-polyploidization, or hybridization between two species that had separated approximately 40 million prior⁴³. As a result, a more apt description for the event would perhaps be a Whole Genome Hybridization (WGH). Of the two species that hybridized during the WGH, one was closely related to the ancestor of the *Zygomycetes*/*Torulospora* (ZT) branch of pre-WGH species, and the other likely separated from the ancestral lineage around the same time as the ancestor of the *Kluyveromyces*/*Lachancea*/*Eremothecium* (KLE) branch (Fig 3.1A).

In the budding yeast lineage, several characteristic phenotypes underwent striking changes following the WGH, including a drastic change in metabolic lifestyle between Crabtree yeast (from post-WGH species) and non-Crabtree yeast (from most pre-WGH yeast species)^{44,45}, as well as changes in the Environmental Stress Response (ESR)^{5,6}. The ESR

is a common gene expression program consisting of hundreds of genes that are either activated or repressed in a variety of stressful conditions such as glucose depletion, osmotic shock, and oxidative stress^{7,46}. Previous studies have demonstrated that much of the ESR is conserved across budding yeast species, and that the level of conservation depends on whether one considers the genes activated during the ESR or those genes that are repressed. The repressed ESR, which contains genes related to ribosomal biogenesis and protein production is more conserved across different yeast lineages, while the activated ESR is more variable^{5,6}. Several of the evolutionary changes that occurred in the activated ESR following the WGH are consistent with changes in metabolic lifestyle between Crabtree and non-Crabtree yeast, including more induction of amino acid and purine biosynthesis genes as well as oxidative phosphorylation genes and less repression of mitochondrial genes in post-WGH species than in pre-WGH species⁵.

The Protein Kinase A (PKA) pathway is one of the primary master regulators of the ESR⁷⁻⁹. Cells with a hyperactive PKA pathway, in addition to being sensitive to stress, also fail to grow on carbon sources that require respiration^{47,48}. While comparative analyses of gene expression between budding yeast species have described broad patterns of conservation and divergence of the genes in the ESR alongside the emergence of the respiro-fermentative lifestyle^{5,6,49}, the precise evolutionary rewiring of the PKA regulon itself has not been clearly delineated. The reason is that most previous studies have used environmental perturbations to elicit and measure transcriptional changes in different species. These perturbations have pleiotropic effects, thus making it difficult to identify the precise transcriptional rewiring that underlies these changes. Precise perturbations to PKA itself are more appropriate for addressing this question unambiguously. One such perturbation is a gatekeeper mutation in the catalytic subunits of PKA that allows precise chemical control of its kinase activity⁵⁰⁻⁵². For PKA in *S. cerevisiae*, this strategy consists of mutating a “gatekeeper” Methionine residue to a less bulky Glycine residue for all three PKA catalytic subunit isoforms (M164G, M147G and M165G for TPK1, TPK2, and TPK3

respectively). The resulting expanded active site still accepts ATP and allows the kinase to phosphorylate its targets with close to wild-type efficiency. However, when a bulky ATP analog is provided to the media, it effectively inhibits any PKA activity. This PKA analog sensitive (AS) allele strategy was used to investigate many properties of PKA signaling in *S. cerevisiae*^{8,9,15,53–55}.

In this work, we employ this AS allele in *S. cerevisiae* and construct an analogous allele in the yeast *Kluyveromyces lactis* (which diverged from *S. cerevisiae* before the WGH) to interrupt PKA signaling in both species and probe the specific evolutionary rewiring of the PKA regulatory program following the WGH. Our data recapitulated many previous findings about the ESR, such as repression of ribosomal proteins in both species, repression of mitochondrial ribosomal proteins in *K. lactis* but not in *S. cerevisiae*, induction of proteins related to mating and meiosis in *K. lactis* but not in *S. cerevisiae*, and induction of genes involved in oxidative stress or generation of metabolites/energy specifically in *S. cerevisiae* but not in *K. lactis*.

Unexpectedly, however, our dataset contained a strikingly large number of WGH ohnologs (at least 91) that showed differential expression under PKA inhibition in *S. cerevisiae*, with one ohnolog being induced under this perturbation and the other with constant or even decreased expression. Furthermore, in this set, on average, the ohnolog that was induced by PKA inhibition had lower basal expression and the uninduced ohnolog had higher basal expression. The phenotype for the shared ortholog in *K. lactis* largely resembled that of the uninduced ohnolog – high basal expression and no induction following PKA inhibition. We explored this observation in depth using publicly available data on gene expression for a range of budding yeast species spanning the WGH^{5,6,16}. Our investigations revealed that, for these differentially expressed ohnologs, the transcriptional response to PKA inhibition of their shared ortholog in the last common ancestor of *S. cerevisiae* and *K. lactis* generally featured low basal expression and low induction, resembling that of the shared

K.lactis orthologs.

To gain insight into the evolutionary trajectories of these ohnologs, we carried out binding site analysis of their gene promoter sequences from a large set of recently published budding yeast genomes¹⁷, focusing on patterns of motifs for Msn2/4, the canonical transcription factors downstream of PKA. These analyses revealed that for some genes, differential expression arose when the ancestral paralogs were in different species prior to the WGH while for others, this divergence is likely to have occurred afterwards. For example, differential expression in response to PKA inhibition for GPM2/3 seems to have evolved long after the WGH, and differential expression for both EGO2/4 and GND1/2 appears to have evolved in the ZT branch prior to the WGH. Interestingly, in the case of GND1/2, a gene conversion may have rendered the phylogenetic history of the protein sequence separate from that of the cis-regulatory region.

In addition to tracing the evolutionary routes of a large part of the PKA regulon, we believe that we present here the first example of divergent regulation that arose in the parent species of the budding yeast before the WGH, and which remains preserved in ohnologs in post-WGH descendants. Given the abundance of allo-polyploid events in diverse species from plants⁵⁶ to vertebrates⁵⁷ this finding may represent a more general avenue for generating novelty in evolution.

3.3 Results

3.3.1 A PKA analog sensitive allele in *K. lactis* allows for comparing the transcriptional response to PKA inhibition with that of *S. cerevisiae*.

Signals transmitting information about stress and carbon source availability alter the cellular level of cAMP which, when in abundance, binds to the regulatory subunit of the PKA heterotetramer, releasing the active subunit to phosphorylate its targets. When PKA is inhibited, the paralogous transcription factors Msn2 and Msn4 are dephosphorylated, translocate into the nucleus and activate a majority of the genes that make up the activated portion of the ESR^{7,11,12,26}. Simultaneously, the transcriptional repressors Tod6 and Dot6 translocate into the nucleus to repress ribosomal biogenesis genes, while Sfp1, which activates ribosomal biogenesis genes, translocates out of the nucleus⁵⁸. The genes regulated by Dot6, Tod6 and Sfp1 are part of the repressed ESR. The response to PKA inhibition also entails a slowdown in growth in *S. cerevisiae*.

In order to compare the response to PKA inhibition between the pre-WGH species *K. lactis* and the post-WGH model species *S. cerevisiae*, we constructed a strain of *K. lactis* in which PKA can be inactivated by addition of the ATP analog 1-NM-PP1. This analog sensitive allele of PKA (PKA-AS) contains a gatekeeper mutation to each of the two PKA regulatory subunits present in that species (M222G for KL.TPK2 and M147G for KL.TPK3). We used a single plasmid CRISPR/Cas9 gene editing strategy⁵⁹, and delivered the Cas9 expression cassette and sgRNA on a universal ARS plasmid that allows for stable expression in a number of budding yeast species⁶⁰. For comparison, we also used *S. cerevisiae* strains containing similar mutations to each of the three PKA subunits¹¹. The PKA-AS alleles were successful in controlling PKA activity in *K. lactis* and *S. cerevisiae*,

as addition of 1-NM-PP1 to these cells stalled growth (Fig 3.7A). Therefore, the role of PKA in inhibiting growth is likely to be conserved between these two species.

The *K. lactis* ortholog of Msn2 and Msn4 (KL.Msn2) has been shown to be important for PKA dependent regulation of mating pathway genes³². In order to directly investigate whether nuclear localization of KL.Msn2 is regulated by PKA as its orthologs, Msn2/4, are in *S. cerevisiae*, we constructed a fluorescently tagged KL.Msn2-RFP in the PKA-AS strain. We used an Msn2 allele in which we mutated residues in the DNA binding domain to ablate DNA binding and minimize the effects of overexpressing the transcription factor¹². For comparison, we also constructed a similar Msn2 reporter in *S. cerevisiae* PKA-AS (SC.Msn2-RFP). Following inhibition of PKA by the drug in *K. lactis* PKA-AS cells, KL.Msn2-RFP strongly located to the nucleus, just as SC.Msn2-RFP did in *S. cerevisiae* (Figs 3.1B, 3.7B). These data clearly indicate that the nuclear localization of Msn2 is regulated by PKA in both species.

3.3.2 The transcriptional responses to PKA inhibition in *S. cerevisiae* and *K. lactis* share many similarities but also exhibit clear functional differences.

To gain a broader understanding of the similarities and differences in the global transcriptional response to PKA inhibition between *S. cerevisiae* and *K. lactis*, we carried out mRNA sequencing for both species in response to saturating amounts of 1-NM-PP1 in exponentially growing PKA-AS cells, measured at 50 minutes after exposure to the drug. We computed the Log Fold Change (LFC) in gene expression using the DESEQ2 algorithm by comparing read counts from yeast treated with drug to read counts from yeast provided only a DMSO control⁶¹. The transcriptional responses assessed using this metric were broadly correlated between the two species (Pearson Correlation = 0.66) (Fig 3.1C),

consistent with the observation that the ESR, and particularly the repressed ESR, is broadly conserved in budding yeasts⁵.

To identify shared and species-specific functional enrichment, we first defined sets of genes that were induced or repressed by PKA inhibition in each species using a threshold that took into account both LFC and p-value (Fig 3.8). Focusing only on genes that had orthologs in both species, we then further subdivided these target sets into subsets that were either activated in both species, repressed in both species, activated in one species but not another, or repressed in one species but not the other. We then quantified enrichment for GO-SLIM terms² for the different gene subsets using Fisher’s exact test (Table 3.1, Fig 3.1C). Go terms enriched in the genes repressed in both species were related to protein production in the ribosome such as rRNA processing ($p=2.38E-87$) and cytoplasmic translation ($p=4.71E-74$), consistent with the slowdown in growth that is seen following PKA inhibition in both species. The genes repressed only in *K.lactis* were enriched for GO-SLIM terms related to mitochondrial translation ($p=8.46E-33$), in agreement with previous observations, and consistent with PKA’s involvement in the shift to respiratory metabolism in *S. cerevisiae*^{5,16,62,63}. The set of genes induced in *K. lactis* but not in *S. cerevisiae* was enriched for meiotic cell cycle control ($p=2.19E-3$) and conjugation ($p=5.86E-3$) consistent with the notion that *K. lactis* incorporates nutritional signals into its decision to undergo meiosis⁶⁴. The set of 151 orthologs induced by PKA inhibition in both species was enriched for carbohydrate metabolic process ($p=2.89E-6$), oligosaccharide metabolic process ($p=4.66E-4$), and vacuole organization (including a number of autophagy related genes) ($p=1.47E-3$), indicating that PKA inhibition triggers a set of conserved metabolic changes. The set of genes that were specifically induced in *S. cerevisiae* but not in *K. lactis* following PKA inhibition was also enriched for carbohydrate metabolic processes ($p=8.56E-6$), but in addition included a significant enrichment for genes involved in generation of precursor metabolites and energy ($p=4.71E-10$), response to oxidative stress ($1.73E-6$), and cellular respiration ($2.21E-5$). This is again consistent with

the role of PKA in initiating respiratory metabolism in *S. cerevisiae*.

3.3.3 Many ohnolog pairs are differentially activated by PKA inhibition in *S. cerevisiae*.

Since *S. cerevisiae* has undergone a WGH, we were curious whether its surviving ohnologs were equally represented in its PKA regulon compared to other genes that didn't maintain their paralogous copy from the WGH. Interestingly, we found that ohnologs were enriched both in the set of genes that were activated by PKA inhibition ($p=2.67\text{E-}10$) and those that were repressed by PKA inhibition ($p=4.88\text{E-}6$), but not in the set of genes whose expression didn't change during PKA inhibition (Table 3.2, Fig 3.1D insets).

To further explore this observation, we plotted the LFC of ohnolog pairs against one another, ordered by their LFC values (Fig 1E). We then searched for ohnolog pairs where one member of the pair was either activated or repressed more than two fold ($LFC > 2.0$ for or $LFC < -2.0$, respectively) with a $\log_{10}(pvalue)$ less than -1.5. We then retained from this set members where the other ohnolog in the pair did not change in the same direction, specifically having LFC less than 1.5 or LFC greater than -1.5 respectively. We also further required that the difference of estimated LFC between the two paralogs be greater than 2.0 in LFC (Fig 3.1E, Table S3). Of the 509 ohnolog pairs in the *S. cerevisiae* genome which had LFC and p-value data for both ohnologs in our dataset, 129 ohnolog pairs had at least one ohnolog repressed and 148 had at least one ohnolog activated according to these criteria. Of the 129 ohnolog pairs that had at least one ohnolog repressed, only 30 pairs (23%) were differentially expressed. Strikingly, of the 148 ohnolog pairs that had at least one ohnolog activated, 91 of them (61%) satisfied these criteria for divergent behavior in response to PKA inhibition pairs. Furthermore, 20 ohnolog pairs had one member activated and the other repressed. Taken together, these data demonstrate

that ohnolog pairs with at least one member activated by PKA inhibition were more likely to have divergent responses than those with at least one member repressed by PKA inhibition.

Functional examination of the different ohnolog pairs showed that those in which both ohnologs were repressed by PKA inhibition were enriched for ribosomal genes. Of the 129 ohnolog pairs that had at least one member repressed, 61 pairs had both members repressed and all but 10 of them had at least one member associated with the GO-slim term “cytoplasmic translation” (Fig 3.1F). By contrast, only 2 of the other 68 ohnolog pairs with one member repressed were associated with this term. These ribosomal ohnolog pairs presumably make up much of the signal that was observed in previous work describing enrichment for ribosomal proteins in retained *S. cerevisiae* ohnologs^{65–67}.

Another prominent difference between the ohnologs repressed by PKA inhibition and those activated by PKA inhibition was in whether their orthologs in *K.lactis* responded similarly to PKA inhibition. Considering only ohnolog pairs that had a *K.lactis* ortholog and expression data in our dataset, there were 122 ohnolog pairs in which one member was repressed by PKA inhibition in *S. cerevisiae*. Of those, 77 (63%) had a *K. lactis* ortholog that was also repressed. By contrast, of the 138 ohnolog pairs in which one member was activated by PKA inhibition in *S. cerevisiae*, only 28 (20%) had *K. lactis* orthologs that were also activated.

Because of the intriguing pattern in the differentially activated ohnologs (which we will simply refer to as DE_{PKA} in the following), we turned our focus to exploring them exclusively in more depth. First, we probed their basal expression because lack of activation for one member of each ohnolog pair in this DE_{PKA} set could correspond to either low or high basal expression. To determine which scenario was more applicable, we examined basal expression of DE_{PKA} genes with and without PKA inhibition, which we quantified using average regularized log counts (rlog) data⁶¹. The basal expression values

of DE_{PKA} genes that had higher LFC values (DE_{PKA} high-LFC ohnologs) had a distribution with a noticeably lower mean than that of all genes in the genome (4.3 vs 5.9 rlog) (Fig 3.2A left panel, Fig 3.2C top panel, Fig 3.8). Furthermore, upon PKA inhibition, the distribution of expression of this group shifted clearly higher than that of all other genes (mean = 7.3 rlog), recapitulating the large LFC for which they were chosen. By contrast, the distribution of basal expression values for the DE_{PKA} genes that had low LFC values (DE_{PKA} low-LFC ohnologs) was higher (mean = 6.9 rlog) and shifted down upon PKA inhibition (mean = 6.1 rlog) (Fig 3.2). Importantly, the distribution of basal expression of the DE_{PKA} high-LFC ohnologs had a clear lower mean than that of the DE_{PKA} low-LFC ohnologs (Fig 3.2A left panel, Fig 3.2C top panel, Fig 3.9), indicating that the difference in LFC between these two sets is the result of low basal expression and strong induction in one case, and a high basal expression that either slightly reduced or unchanged by PKA inhibition in the other case. *K. lactis* contains orthologs for 87 of the 91 ohnolog pairs in the DE_{PKA} set. As a group, these orthologs had a similar phenotype to that of the DE_{PKA} low-LFC ohnologs, displaying high basal expression (Fig 3.2A right panel, 3.2C lower panel) and little change in LFC under PKA inhibition (Fig 3.2A right panel, 3.2B right column).

Taken together, our data and analyses so far reveal a class of ohnologs, amounting to about 1/6 of all the retained ohnolog pairs, or about 3% of all genes in *S. cerevisiae*, which show differential expression under PKA inhibition with one member of an ohnolog pair exhibiting high induction and the other showing either no change or repression. On average, the PKA insensitive ohnolog exhibits high basal expression. The *K. lactis* orthologs of these differentially expressed ohnologs have a similar lack of induction and high basal expression. This observation posed the hypothesis that the ancestral phenotype for many of these ohnolog pairs might have been one with high basal expression and no susceptibility to PKA.

3.3.4 High basal expression and low LFC under PKA inhibition seems to be the ancestral state for most DE_{PKA} genes.

Without gene expression measurements following PKA inhibition in a wider phylogenetic range of extant species, testing whether low LFC in response to PKA inhibition was the ancestral state of the DE_{PKA} genes cannot be done directly. However, we can explore it meaningfully across a range of 15 budding yeast species spanning the WGH by capitalizing on available gene expression datasets collected in response to growth to saturation and various stress conditions known to involve the PKA pathway^{5,6}. To extract the conditions that most resemble PKA inhibition, we compared expression changes in *S. cerevisiae* and *K. lactis* contained in these datasets to our PKA inhibition data in *S. cerevisiae* and *K. lactis*. This comparison identified five conditions (heat shock at 30 and 45 minutes, Diauxic Shift, Post Diauxic Shift, and Plateau) that were strongly correlated with PKA inhibition in both species (Pearson correlation $R \geq 0.65$ with *K. lactis* PKA inhibition and $R \geq 0.75$ with *S. cerevisiae* PKA inhibition) (Fig 3.10). We adopted LFC reported for these 5 conditions (normalized per Materials and Methods) as surrogates for PKA inhibition in all species.

We curated the orthologs of the 91 DE_{PKA} ohnolog pairs in the various species and compared their gene expression values in the 5 PKA-related stress conditions. The different members of the ohnolog pairs in the DE_{PKA} set generally showed the same phenotype in *S. cerevisiae* as in our PKA inhibition dataset —high LFC for one member of the pair and low LFC in the other (Fig 3.3A). This is expected given our criteria for selecting these stress conditions. In the post-WGH species closely related to *S. cerevisiae*, the phenotype was also conserved, with syntenic orthologs of the high-LFC ohnologs having higher LFC, and those of the low-LFC ohnologs having lower LFC on average. The separation between the ohnolog pairs declined in more distantly related Post-WGH species. In *V. polyspora*, the most distantly related Post-WGH species, the orthologs of the low and high LFC

ohnologs both had low LFC in stress conditions (Figs 3.3A and 3.11). In interpreting this result it is important to note that assignment of syntenic orthologs becomes difficult for distantly related post-WGH species such as *S. cerevisiae* and *V. polyspora*. For pre-WGH species, the shared orthologs of the DE_{PKA} ohnolog pairs have low LFC under these PKA-related stress conditions (Fig 3.3A), consistent with our hypothesis that low LFC is generally the ancestral response to PKA inhibition for DE_{PKA} ohnolog pairs.

To explore basal expression of orthologs of the DE_{PKA} genes in these species, we turned to a different dataset that measured gene expression during exponential growth on rich media using microarray data for a range of budding yeasts¹⁶. After appropriate normalization (see Materials and Methods), we checked concordance of these data to rlog estimates from our RNA seq count data for *S. cerevisiae* and *K. lactis* strains under control conditions and observed high correlation despite the different methods by which gene expression data was collected (Pearson Correlation $R=0.69$ and 0.60 for *S. cerevisiae* and *K. lactis* respectively) (Fig 3.12A). Reassuringly, there was high correlation between both datasets for the DE_{PKA} genes ($R=0.67$ for the low-LFC ohnologs and 0.70 for the high-LFC ohnologs) as well as for their orthologs in *K. lactis* ($R=0.59$) (Fig 3.12B), instilling confidence that the various sets are coherent in their content and can be used for cross-comparison.

With these data in hand, we could compare basal expression of DE_{PKA} genes and their syntenic orthologs across species. For the post-WGH species, the syntenic orthologs of the low-LFC DE_{PKA} ohnologs generally had a higher basal expression while those of the high-LFC DE_{PKA} ohnologs generally had lower basal expression for species more closely related to *S. cerevisiae* and higher basal expression for species more distantly related. For the pre-WGH species, the basal expression of the shared ortholog of a given DE_{PKA} ohnolog pair was typically high, similar to that of the low-LFC DE_{PKA} ohnolog in *S. cerevisiae* (Figs 3.3B, 3.137). All conclusions also held when we redefined the ohnolog pairs of interest based on the 5 stress conditions rather than on the PKA-inhibition dataset (Figs

3.14,3.15, 3.16). Taken together, the evidence supports our hypothesis that the basal expression level of the shared ancestors of the DE_{PKA} ohnolog pairs was, on average, high.

The use of stress conditions also allowed us to define differentially expressed ohnolog pairs from the perspective of post-WGH species other than *S. cerevisiae*, therefore enabling us to test whether our conclusions about the state of the ancestral phenotype is sensitive to the species used as a reference. Towards this, we defined the set of differentially expressed ohnologs in *N. castellii* and *V. polyspora* using gene expression for the five PKA-related conditions (see Materials and Methods). Similar to when *S. cerevisiae* was used as a reference, orthologs in pre-WGH species of these ohnologs also had low LFC values (for ohnologs defined using *N. castellii* and *V. polyspora*) and high basal expression (for ohnologs defined using *N. castellii*) (Figs 3.15, 3.16).

Taken together, these data strongly support the notion that the ancestral state of ohnologs that are differentially expressed in response to PKA was characterized by high basal expression and insensitivity to PKA. However, because the pre-WGH species for which stress response data was present all diverged from the *S. cerevisiae* lineage at the same time or earlier than *K. lactis*, these data do not indicate whether the derived phenotypes (high LFC in response to PKA inhibition and low basal expression) for any particular ohnolog arose before the WGH in the lineage leading to the ZT branch, or after the WGH (Fig 3.3C). To be made with certainty, this inference would require knowledge of gene expression in the *Zygomycetes/Torulospora* (ZT) branch in response to PKA related stresses. Since no such data exist, we next turned to exploring whether the promoters of the DE_{PKA} genes have a bioinformatic signal that might correlate with these phenotypes, capitalizing on the wealth of sequenced genomes in the budding yeast subphylum to formulate hypotheses about their evolutionary trajectory¹⁷.

3.3.5 The STRE is enriched in the promoters of the genes induced by PKA inhibition in *S. cerevisiae*, and *K. lactis*.

To generate a framework for evaluating the characteristics of DE_{PKA} gene promoters, we first investigated bioinformatic signals associated with the promoters of all genes activated by PKA inhibition. To do so, we used the DREME algorithm (see Materials and Methods) to identify short ungapped motifs that are enriched in comparison to a background set of promoters in *S. cerevisiae* (Bailey, 2011). Comparing the promoters of all genes activated under PKA inhibition in *S. cerevisiae* against the promoters of all *S. cerevisiae* genes, we identified the Stress Response Element (STRE, CCCCT), the sequence for the binding site for Msn2 and Msn4, as the most heavily enriched motif (E-value 3.3e-40) (Fig 3.4A). This is consistent with previous findings, implicating Msn2 and Msn4 as prominent targets downstream of PKA^{10,68}. Four of the other five motifs enriched in the promoters of genes activated by PKA inhibition were similar to the STRE or the Post Diauxic Shift element (PDS) motif (T(A/T)AGGGAT) which is itself structurally similar to the STRE⁶⁹, while others resembled the TATA-box (TATA(A/T)A(A/T)(A/G)) (E-value 1.3e-3) which is known to be enriched in stress responsive promoters⁷⁰.

Next, we asked how the number and locations of STRE and TATA-box motifs correlated with a gene being responsive to PKA inhibition. Promoters of genes activated by PKA had a larger probability of containing one or more STREs relative to all promoters in *S. cerevisiae* (75.2% v.s. 43.9%, p-value 1.5e-16). They also had a notable increase in the average number of STREs per promoter (1.32 v.s. 0.62) (Fig 3.4B). Furthermore, the location of the STREs in the promoters of the genes induced by PKA inhibition had a unimodal distribution with 64.7% of STRE sites found between 100 and 400 base pairs, as opposed to an expectation of 42.9% from a uniform distribution and a 46.1% value when the distribution of STRE locations is compiled for all promoters in the genome (Fig 3.17).

Analysis of TATA-box motifs showed a similar pattern (70.7% of motifs found between the first 100 and 400 base pairs versus 57.7% for all genes, p-value $2.3e-3$) (Fig 3.18A). Finally, TATA box and STRE motifs are more likely to occur together in promoters of genes activated by PKA inhibition than in all genes (43% of promoters v.s. 18%), as expected from the increased percentages of both STREs and TATA boxes in genes activated by PKA inhibition (Fig 3.19).

Probing the bioinformatic characteristics of promoters of genes that are responsive to PKA inhibition in *K. lactis* revealed that those also had more STREs on average compared to the promoters of all genes. However, the top hit for promoters of genes activated by PKA inhibition in *K. lactis* was a motif whose Position Specific Scoring Matrix (PSSM) would be satisfied by an STRE but was closer to a PDS (E-value $4.4e-19$) (Fig 3.4A).

Furthermore, the bioinformatic signal for the number of STREs and their location was weaker in *K. lactis* than in *S. cerevisiae* (47.7% of promoters with 1 or more STRE in the promoter in PKA activated genes v.s. 34.5% in all genes, $p=1.6e-4$) (Figs 3.4B and 3.17). In *K. lactis*, as in *S. cerevisiae*, many more promoters of genes activated by PKA had TATA boxes than all promoters in the genome (70.8% v.s. 54.1% with 1 or more TATA box in the promoter, $p=3.8e-4$) (Fig 3.18B).

Taken together, these results indicate that the presence of STRE and TATA box motifs in promoters might be a useful proxy for activation by PKA inhibition across budding yeast species. We therefore looked for these features in the promoters of DE_{PKA} ohnologs and their *K. lactis* orthologs to see whether the presence of STRE and TATA box motifs correlated with the response to PKA we observed in our experiments. The distributions for the number and location of STREs for the DE_{PKA} high-LFC ohnologs closely resembled those of the promoters for genes activated by PKA inhibition. However, importantly, the distribution for the DE_{PKA} low-LFC ohnologs resembled that of promoters of all *S. cerevisiae* genes (Figs 3.4B, 3.17, and 3.19). Furthermore, the distribution of the number of

STREs in the promoters of *K. lactis* orthologs of the DE_{PKA} genes was close to that of all genes in *K. lactis* (33% v.s. 34.5% had one or more STRE in the promoter), unsurprising in light of the fact that these genes had low LFC in response to PKA inhibition.

The TATA-box distributions in the promoters of the DE_{PKA} high-LFC ohnologs were also increased relative to those of the promoters of all genes. However, unlike for the STRE, this enrichment was also present for DE_{PKA} low-LFC ohnologs and the *K. lactis* orthologs of the DE_{PKA} genes (Figs S12 and S13). Based on that observation, we reasoned that, at least in the context of the DE_{PKA} genes and their orthologs, the TATA box was not linked strongly enough to induction following PKA inhibition, and was likely to be an ambiguous evolutionary signal. Therefore, we focused instead on the presence of STREs as a bioinformatic proxy for gene induction in response to PKA inhibition.

3.3.6 Analysis of STREs in the promoters of DE_{PKA} orthologs across species revealed enrichment of STREs in the ZT branch.

To investigate whether the relationship between activation by PKA inhibition and STRE presence in the promoter holds in other budding yeast species, we revisited the cross-species gene expression datasets from^{5,6} and scored STRE enrichment in the promoters of genes activated under conditions correlated with PKA activation in *S. cerevisiae* and *K. lactis* (Fig 3.4C). With the exception of heat stress in *K. lactis* and diauxic shift in *L. waltii* and *L. kluyverii*, the promoters of the genes activated under these stresses had more STREs than all genes in that species (pval<0.05 using Fisher’s Exact test) although the overall background level of STREs and the level of enrichment varied widely between species. These results suggest that there is a relationship between the presence of STREs in the promoter and gene induction under PKA inhibition across

budding yeast that can be detected bioinformatically.

We therefore quantified STREs in the promoters for orthologs of DE_{PKA} genes to search for examples in which the STRE conservation patterns might provide some insight into the evolution of gene responsiveness under conditions of PKA inhibition. We analyzed promoters in 32 pre-WGH and 12 post WGH species, including species for which no stress or PKA inhibition gene expression data exists (e.g. species in the ZT branch). For this analysis, we only focused on the DE_{PKA} ohnolog pairs that in *S. cerevisiae* had at least one STRE in the promoter of the high-LFC member. We discarded genes with high-LFC but no STRE in the promoter, reasoning that they are induced through a mechanism that doesn't require the STRE (for example through a transcription factor besides Msn2/4 downstream of PKA), and thus the conservation of the STRE would not be the most relevant bioinformatic signal. We also removed pairs that had more than one ortholog in more than 8 pre WGH species which included ohnolog pairs that were present as duplicates before the WGH (such as the hexose transporters) and genes that may have undergone a separate ancient duplication which would complicate our analysis (such as SNF3/RGT2). Finally we removed pairs that had that had no orthologs in pre-WGH species (such as USV1/RGM1) (see Materials and Methods).

This amounted to a total of 60 ohnolog pairs (see Materials and Methods). Exploration of the promoters of the orthologs of the high-LFC ohnologs in this set showed an enrichment of STREs in post WGH species (Figs 3.5A, 3.20 red bars, $p < 0.05$). This enrichment was not present in the low-LFC ohnologs (Figs 3.5A, 3.21 blue bars). This pattern was more prominent for species closely related to *S. cerevisiae*, with increased variability for species more distantly related and no enrichment for *V. polyspora*, the most distant post-WGH species we analyzed. *C. glabrata* was an exception to this pattern with enrichment for STREs in both the syntentic orthologs of the low-LFC and high-LFC ohnologs. Further analysis of the promoters of the pre-WGH orthologs of this subset of DE_{PKA} ohnolog pairs

revealed more enrichment for STREs in the ZT branch (7 out of 11 species) than in the KLE branch (only 6 out of 21 species) (Figs 3.5A, 3.20). In the outgroup species, there was a mixed picture across species, with 3 out of 9 showing enrichment for STREs. Despite the fact that these data reflected the average of 60 separate sets of orthologs with their own idiosyncratic evolutionary histories, the evidence hints that some of the differential expression we see in DE_{PKA} ohnolog pairs may have arisen in the promoters of pre-WGH orthologs in the ZT branch prior to the WGH, while some may have arisen following the WGH in the promoter of the ancestor of the high-LFC ohnolog. We explored potential specific examples of both scenarios.

3.3.7 EGO2/4 provides an example where gain of STREs in the ZT branch may have given rise to PKA dependence which was conserved in syntenic orthologs following the WGH.

To identify specific promoters with a clear evolutionary signal, we clustered the subset of 60 DE_{PKA} ohnolog pairs described above based on the numbers of STREs in the promoters of their orthologs across 32 ZT and KLE branch species and 9 outgroups. As expected, there is much variation given that the STRE is only a partial surrogate of PKA dependence. However, a few clusters highlight possible examples of WGH paralogs acquiring differential induction under PKA inhibition via distinct evolutionary trajectories (Figs 3.5, 3.21). Specifically, one cluster of 7 ohnolog pairs was characterized by high numbers of STREs in the ZT branch and relatively fewer STREs in the KLE branch. This pattern would be expected if the motif arose in the ZT branch after the two branches split but before the WGH. EGO2 and EGO4 provide a clear illustrative example of this evolutionary trajectory.

EGO2 and EGO4 in *S. cerevisiae* are short (75 and 98 residues respectively) proteins that

have no clear orthologs outside of the *Saccharomycetacea*, a group which includes the KLE, ZT and post-WGH yeast species. EGO2 is expressed at intermediate levels (6.09 ± 0.12 rlog counts) during exponential growth and does not appreciably change under PKA inhibition (0.41 ± 0.37 fold) (Fig 3.6A inset). EGO4 has low expression during exponential growth (4.12 ± 0.04 rlog) and is highly induced (7.16 ± 0.38 fold) under PKA inhibition. EGO4 is also an Msn2/4 target because its activation following PKA inhibition in an Msn2/4 deletion strain decreases substantially (to 1.45 ± 0.60 fold) (Fig 3.22). It did not meet our threshold to be a target of PKA inhibition in *K. lactis*, and the LFC following PKA inhibition only changed from 1.06 to 0.70 after deletion of Kl. Msn2. EGO2 was recently shown to be part of the EGO complex which is important for recruiting Gtr GTPases to the vacuolar membrane for TORC1 signaling, and may be a structural homolog of LAMTOR5 in mammalian cells⁷¹. While a deletion of EGO2 results in sensitivity to Rapamycin, Δ EGO4 mutants show wild type levels of resistance to rapamycin. Aside from increasing sensitivity to rapamycin in a Δ GTR1 background, the biological function of EGO4 and its possible role in TORC1 signaling remain unclear. To explore the evolutionary history of STRE binding sites in EGO4 and EGO2, we used protein sequence similarity to construct a phylogenetic tree for orthologous yeast proteins and traced binding sites in promoters associated with their coding sequences. The structure of the tree suggested that the protein sequence likely diverged after the ZT and KLE branch split and prior to the WGH, because one branch of post-WGH ohnologs (the one containing EGO4) was more closely related to orthologs in the ZT branch than it was to the other branch of post-WGH ohnologs (the one containing EGO2) (Fig 3.6A). All the promoters for the ZT branch had at least 1 STRE within 700 bases of the transcription start site, and a TATA box within 300bp of the transcription start site. The location of these motifs was in some cases conserved across species. This pattern was preserved for the syntenic orthologs of EGO4. The pattern was generally not present for the promoters of the EGO2 post-WGH orthologs, nor for the promoters of the KLE orthologs, in which

STRE and TATA motifs were scattered more sporadically. The phylogenetic distribution of this pattern paints a picture in which STREs arose upstream of the ancestral EGO2/4 ortholog in the ZT branch prior to the WGH and were then conserved. The promoter of EGO4 is likely to have descended from the WGH parental strain more closely related to the ZT branch containing this conserved STRE, potentially rendering it inducible by PKA inhibition. In this scenario, the promoter for the ancestor of EGO2 would have descended from the WGH parental strain more closely related to the KLE branch, would have had no STREs, and would not have been induced by PKA inhibition.

3.3.8 The STRE in the promoter for GND2 may have arisen in the ZT branch (similar to EGO2/4) but its evolutionary history is obscured by gene conversion.

GND1/2 are an ohnolog pair in the same cluster as EGO2/4 (Figure 5B). GND1/2 are two of the 12 enzymes in the Pentose Phosphate pathway⁷². Interestingly, our DE_{PKA} set of differentially expressed ohnologs contains three additional ohnolog pairs that are part of the Pentose Phosphate Pathway. As with EGO2/4, most (10/12) of the promoters of the ZT branch orthologs of GND1/2 contain at least one STRE in the 700bp upstream of the start codon and a TATA box within 300bp (Fig 3.6B). Only 8/21 of the promoters of the KLE branch orthologs have these features, suggesting that the STRE may have arisen in the ZT branch prior to the WGH. The interpretation following the WGH, however, is less clear than in the case of EGO2/4.

For *S. cerevisiae* and the other sensu stricto species (i.e. with the *Saccharomyces* genus name such as *S. mikatae*) the protein sequences for GND1 and GND2 are more closely related to one another than either are to orthologs from other Pre-WGH species. This suggests that there was a gene conversion in which all or part of one of the proteins was

overwritten through homologous recombination in the common ancestor of the *sensu stricto* species. Gene conversions are known to occur frequently after interspecies hybridization events, specifically after the WGH in budding yeast^{43,73}. In the promoters of the *sensu stricto* orthologs more closely related to GND2, there are two conserved STREs upstream of a conserved TATA box. These could have arisen following the gene conversion, however it is also possible that these regulatory regions were conserved from before the WGH and were present in the promoter prior to the gene conversion. In this way, the phylogeny of the regulatory region may be separate from that of the protein sequence.

3.3.9 GPM2/3 illustrates a case in which STREs might have been added following the WGH in the *sensu stricto* lineage.

A separate cluster in Figure 3.5B was characterized by low numbers of STREs in the promoters of orthologs from both the ZT branch and the KLE branch. Assuming that the presence of STREs is a major driver of expression upon PKA inhibition for these ortholog pairs, then the STREs and ensuing differential expression would have arisen in these pairs following the WGH. GPM2/3 provides an example of this situation (Fig 3.6C).

GPM2 and GPM3 were identified in yeast as homologs to the phosphoglycerate mutase enzyme, GPM1, which is part of the glycolysis and gluconeogenesis pathways. Both genes are only able to rescue the growth defects of a GPM1 deletion strain when they are overexpressed at high levels⁷⁴. GPM2 has been shown to be important for growth under the respiratory carbon sources glycerol and sorbitol, and seems to play a role in the maintenance of cellular membrane stability⁷⁵. A specific function for GPM3 has not yet been identified.

In our dataset, GPM2 mRNA levels increased substantially under PKA inhibition (5.16 ± 0.52 LFC), while GPM3 maintained a stable (albeit low) expression level

(3.79 ± 0.87 rlog counts, -0.26 ± 0.75 LFC). While most pre-WGH species contained a shared ortholog for GPM2/3, in the post-WGH species, it was only retained in duplicate in the sensu stricto species. A conserved STRE sequence seems to have arisen in the promoter sequence of the common ancestor of the syntenic orthologs of GPM2 from the sensu stricto clade. This motif was presumably part of what gave rise to differential expression between the two paralogs in that clade (Fig 3.6C). However, based only on this motif data, we cannot say whether these STREs arose in the regulatory region surrounding the gene prior to or after the gene conversion.

3.4 Discussion

Gene duplication is a major driver of innovation in evolution, and our goal in this study was to better understand the mechanisms that can drive this evolutionary innovation in the special case of an allo-polyploidization or Whole Genome Hybridization (WGH). One of the most well studied examples of a WGH occurred in budding yeast in the branch leading to the canonical single celled eukaryote, *S. cerevisiae*, and the WGH has been implicated as a key factor that enabled the change in metabolic lifestyle from respiratory (or Crabtree negative) to respiro-fermentative (Crabtree positive) that occurred at the same time as the WGH in that lineage⁴⁵. It is known that the paralogs generated from the yeast WGH (ohnologs) are often differentially expressed under stress conditions. However, the pleiotropic nature of environmental stress makes it difficult to parse the contribution of different interacting stress response pathways to differential expression. By narrowing our study to the PKA pathway, a conserved master regulator of the Environmental Stress Response, and capitalizing on ways to perturb it specifically, we could meaningfully explore how the signal from PKA was delivered to ohnologs that are differentially expressed. Additionally, in comparing the transcriptional response to PKA inhibition between the pre-WGH species

K. lactis and the post-WGH species *S. cerevisiae*, direct control of PKA ensured that changes we observed were the result of changes downstream of this particular pathway rather than changes in the stress sensing machinery and the crosstalk between pathways.

Our data established a strong enrichment for ohnologs from the yeast WGH in the set of genes activated by PKA inhibition in *S. cerevisiae* and revealed that many of these ohnologs had differential expression, with one member of the pair showing high induction by PKA inhibition and the other remaining insensitive to PKA, which we denoted DE_{PKA} ohnolog pairs. This set amounts to a large proportion (almost 1/6) of the retained ohnologs in *S. cerevisiae*. Our investigation of the transcriptome of *K. lactis* in response to PKA inhibition, and our further analysis of publicly available gene expression data from a number of budding yeast species in response to stress conditions that approximated PKA inhibition revealed that the orthologs of these differentially expressed ohnologs in various WGH-species had low induction and high basal expression. This suggested that insensitivity to PKA is likely to be the ancestral state for many of these differentially expressed ohnologs^{5,6,16}. To explore the evolutionary timeline for the emergence of PKA dependence for the ohnologs that were induced by PKA inhibition, we looked for bioinformatic signals in the promoters of differentially expressed ohnologs and their orthologs across over 70 sequenced budding yeast genomes. We found compelling evidence that the STRE binding site, which is strongly linked to activation by Msn2/4 under PKA inhibition, was gained in the ZT branch prior to the WGH in some cases and arose following the WGH in others. The examples of gene promoters we further analyzed as examples of these scenarios highlighted additional interesting evolutionary issues.

Given the immense challenge faced by natural selection to maintain selection pressure for redundancy, understanding how duplicated genes are maintained is one of the important questions in understanding evolution. Many hypotheses have been put forward, including the “balance hypothesis”⁶⁶, which posits that paralogous genes that are dosage dependent

(i.e. members of essential complexes) are likely to be retained following a WGD or WGH, since the loss of one copy would cause a substantial fitness defect. This is likely to be the case for the large contingent of ribosomal ohnolog pairs that are simultaneously repressed under PKA inhibition in our data. Other evolutionary forces that work to generate functional divergence in paralogs include escape from adaptive conflict, sub-functionalization, and neofunctionalization^{34,35}. In an autopolyploidization event, or WGD, however, one still must explain how such mechanisms are deployed. Because in a WGD the two copies of a gene are initially identical, one must invoke very strong selection for advantageous functional divergence or easily accessible mutations that provide functional divergence in order for the divergence to arise before deleterious mutations have a chance to break one of the two paralogs.

An allopolyploidization, or WGH, on the other hand provides a simpler explanation for the origin of functional divergence in paralogous genes. Two paralogous genes might have functionally diverged during the time they were separated in different lineages, each adapting to its own specific environment. We envision this to be the case for EGO2/4 genes, which arrived in *S. cerevisiae* with different regulation (one downstream of PKA and one not) and different protein phylogeny. GPM2/3 may also have experienced functional divergence prior to the WGH, but it appears that differential regulation in response to PKA (in the form of STREs in the promoter) did not arise until afterwards. However, this simple story is often greatly complicated by the occurrence of gene conversion following WGH events when homologous genes overwrite all or part of their paralogs either soon after the diploidization or in the following millenia⁷³. It has been proposed that the functionally divergent ohnolog pair GDH1 and GDH3 experienced a gene conversion following the WGH⁷⁶. Our investigations of GND1/2 indicate that they might constitute another example of gene conversion.

Understanding the implications of gene conversion following a hybridization (versus a

duplication) event is therefore crucial. In a hybridization event, while the protein sequence is homogenized through gene conversion, divergent regulation is likely to remain in the form of different cis-regulatory elements unaffected by the gene conversion in the promoters of the paralog pairs. A corollary of that is the idea that a WGH can provide an environment in which functional divergence could evolve as different ohnologs gain mutations that either enhance their function in conditions under which they are expressed or degrade functions required for conditions under which they are not. This may have been the case for GND1/2. The same reasoning applies if cis-regulatory sequences appear de-novo in the promoter of one of two duplicated ohnologs, causing differential expression and paving the way for further functional specialization.

The final confirmation of these insights however awaits measurements of gene expression under direct inhibition of PKA in a range of budding yeast species, including important members of the ZT branch. Such measurements will illuminate how much divergent expression was present in the ancestral hybrid, and how much arose de-novo in the various post WGH lineages. Another fruitful direction in exploring the evolutionary trajectories we hypothesize would be to perturb cells further down the PKA pathway at the level of the transcription factor, Msn2, to avoid the web of crosstalk that might exist downstream of PKA. This is technically challenging for Msn2 because its activity is controlled more by its nuclear localization than by its expression, however technology to optogenetically induce nuclear localization for transcription factors such as Msn2 has recently been demonstrated and might be possible to port into other budding yeast species⁷⁷.

Despite our use of a bioinformatic signal with clear weaknesses, including a relatively low information content (just 5 bases) and in imperfect correlation between the signal (the STRE) and the biological phenotype it was meant to detect (induction in response to PKA inhibition), we were still able to leverage the high density of sequenced species in budding yeast to draw evolutionary conclusions¹⁷, illustrating the usefulness of the budding yeast as

a model phylum for studying evolution³. The rest of the known budding yeast will soon be sequenced completing the goals of the y1000plus project¹⁹ and that set of sequences will represent all the evidence that remains of hundreds of millions of years of regulatory evolution for those species. Our ability to extract insights into regulatory evolution from that data awaits a better understanding at a few key links in the chain connecting the pathways that sense environmental changes (such as PKA) to gene expression. Breakthroughs at a few levels might soon yield the technology necessary to read bioinformatic regulatory signals in these yeast at a much greater level of precision. A screen of pathway perturbations linked to high throughput gene expression readout, such as single colony RNA-seq²³, combined with machine learning approaches would make great strides towards identifying such a bioinformatic regulatory signal for the PKA pathway. This would be strengthened by a tighter link between pathway mutants and transcription factor binding via traditional assays such as ChIP-seq and complementary methods such as the calling card method⁷⁸. It would also be strengthened by incorporating the fuller understanding of the links between transcription factors and the DNA sequences they bind, both in vitro and in vivo^{79,80}. A combination of all such methods will greatly accelerate our understanding of the complex evolutionary trajectories of fungal genes.

3.5 Figures

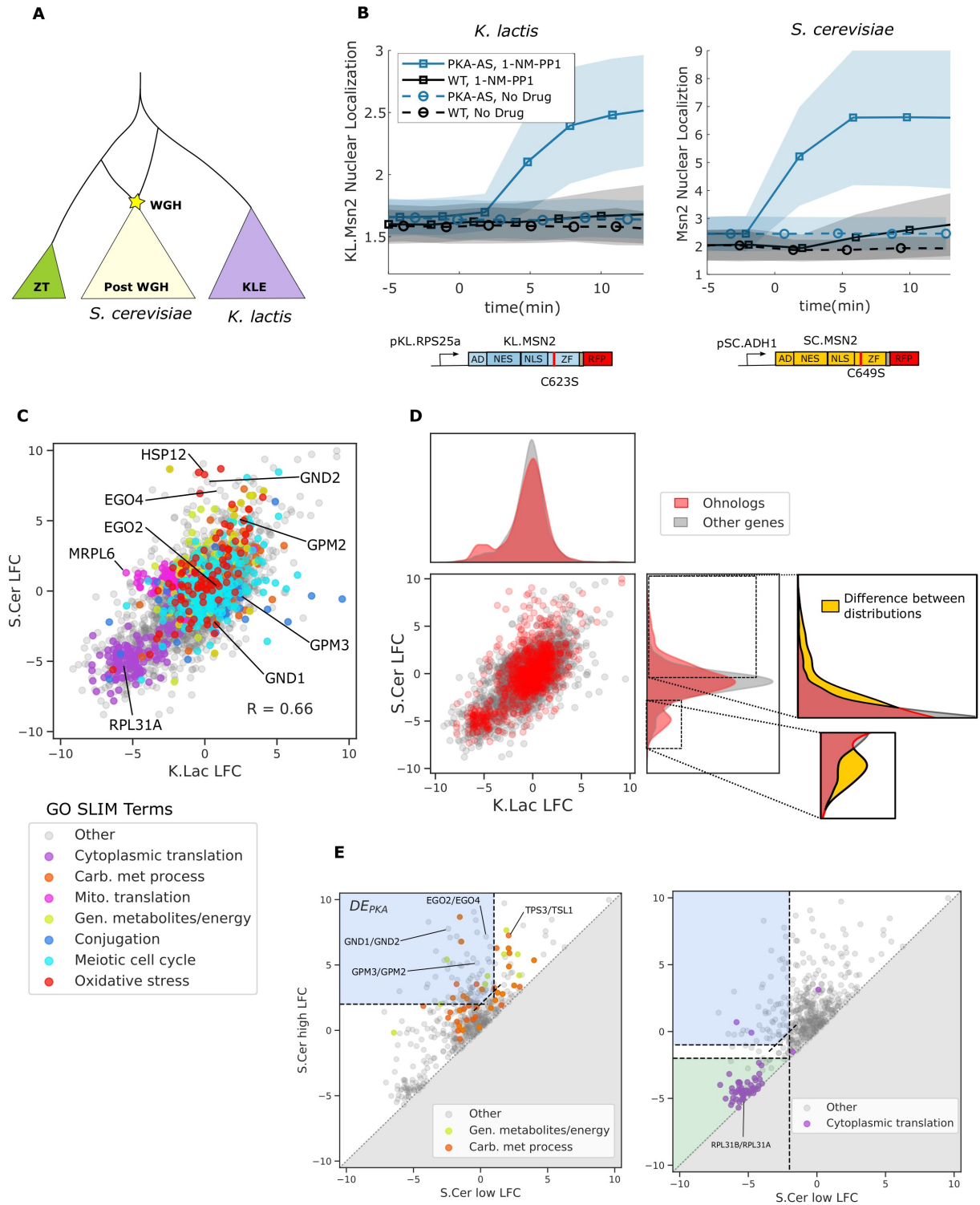


Figure 3.1: **PKA inhibition elicits a broadly similar transcriptional response in *S. cerevisiae* and *K. lactis*, but there are important differences. Ohnologs are enriched in the genes repressed by PKA inhibition and in the genes activated by PKA inhibition, where they are often differentially expressed.** (A) Simplified schematic depicting the budding yeast Whole Genome Hybridization (WGH). The Kluyveromyces/Lachancea/Eremothecium (KLE) branch is shaded purple, and the Zygosaccharomyces/Torulaspora (ZT) branch is shaded green. (B) Msn2 nuclear localization in *K. lactis* (left) and *S. cerevisiae* (right) for WT and PKA-AS strains following the addition of control media or $4\mu M$ 1-NM-PP1. Solid line represents the mean and the shaded area represents the standard deviation for at least 27 single cell measurements in *K. lactis* and at least 84 single cell measurements in *S. cerevisiae*. Diagrams of the KL.Msn2 and SC.Msn2 nuclear localization reporters are shown. Red line indicates the mutation that ablates DNA binding. (C) Log Fold Change (LFC) comparing RNA sequencing data collected by inhibiting PKA-AS with $3\mu M$ 1-NMPP1 versus DMSO. Data is collected after 50 min in *S. cerevisiae* (y-axis) and *K. lactis* (x-axis) following administration of the drug. LFC values are only shown for genes that had orthologs in both species. Genes associated with representative GO SLIM terms that were enriched in analysis are color coded (Table 3.1). (D) Data are as in (C) but with ohnologs highlighted in red and all other genes colored grey. The kernel density estimates of the distribution of LFC for ohnologs in *S. cerevisiae* and for their *K. lactis* orthologs are shown to the right and above the scatter plot respectively. The distribution of LFC for all genes is shown in grey for comparison in both plots. Regions of the distribution highlighting the enrichment of ohnologs in the genes induced by PKA inhibition and repressed by PKA inhibition v.s. all genes are shown with the difference between distributions in yellow. (E) Comparison of LFC for ohnolog pairs in *S. cerevisiae* are shown in both plots. The smallest LFC value for each ohnolog pair is plotted on the x-axis and the largest LFC value is plotted on the y-axis. The data points are colored by selected GO-terms. The blue shading for the left plot indicates the LFC criteria for selecting differentially expressed ohnologs in which one member of the pair is induced by PKA inhibition (DE_{PKA}). The blue shading for the right plot indicates LFC criteria for selecting differentially expressed ohnologs in which one member of the pair is repressed by PKA inhibition. The green shading in the right plot indicates ohnolog pairs in which both genes are repressed by PKA inhibition.

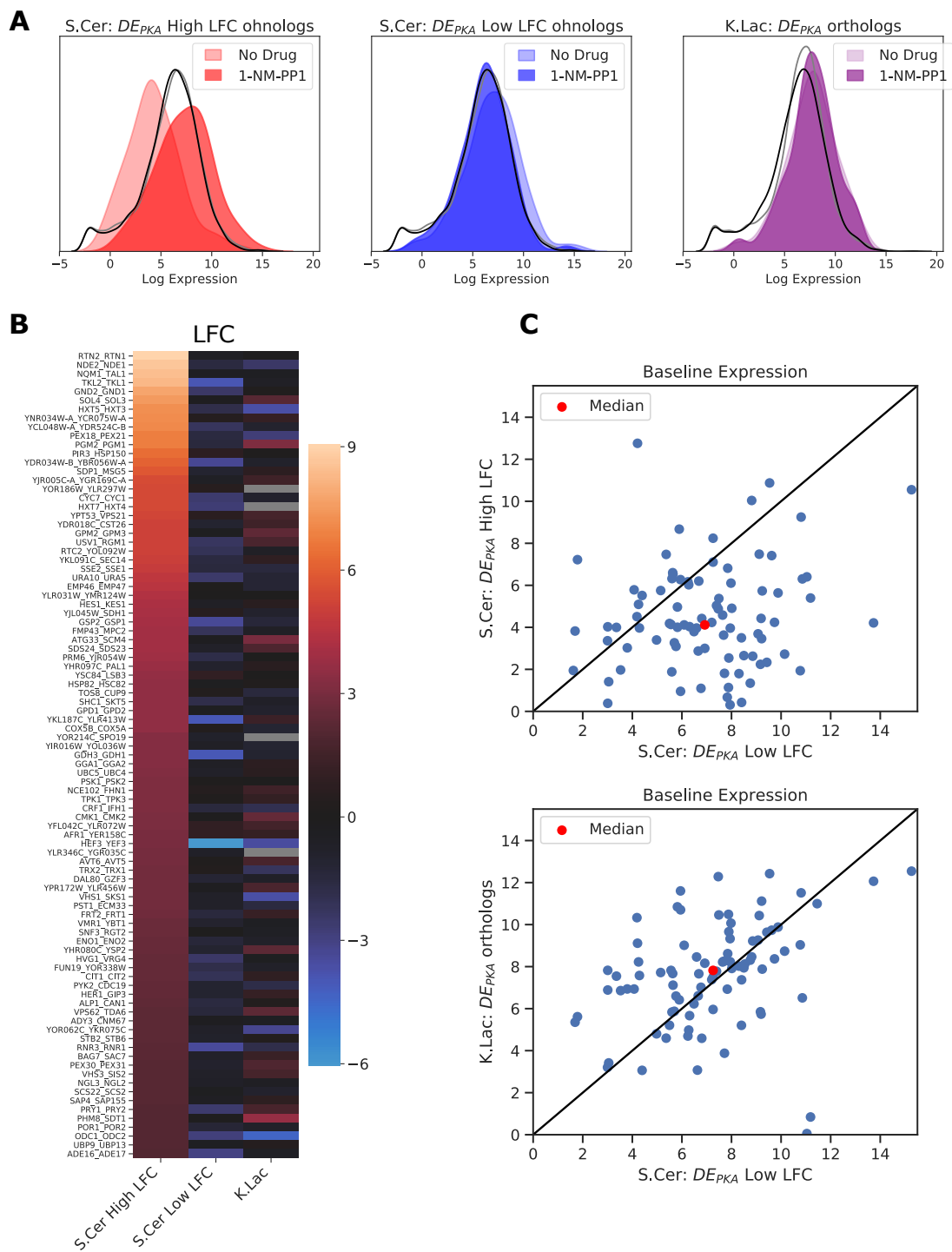


Figure 3.2: **The high-LFC ohnolog from DE_{PKA} tends to have lower average basal expression in *S. cerevisiae*, while the low-LFC ohnolog and their shared ortholog in *K. lactis* tend to have higher average basal expression.** (A) Kernel density estimates for average rlog data DE_{PKA} high and low LFC ohnologs are plotted without treatment (light shading) and with treatment (dark shading) with $3\mu M$ NMPP1. Background expression for all *S. cerevisiae* (excluding dubious orfs) and all *K. lactis* genes with orthologs in *S. cerevisiae* is also shown (grey and black lines, respectively). (B) LFC for low-LFC ohnologs (first column) and high LFC ohnologs (middle column) from DE_{PKA} ohnologs in *S. cerevisiae* as well as their *K. lactis* orthologs (last column). Rows are sorted from highest to lowest LFC for the high LFC-ohnolog. (C) Plot of baseline expression (rlog normalized estimate of counts for no-drug samples) for DE_{PKA} low-LFC ohnologs from *S. cerevisiae* (x-axis) v.s. DE_{PKA} high-LFC ohnolog (y-axis, top panel) and DE_{PKA} orthologs from *K. lactis* (y-axis, bottom panel). Median values are shown in red.

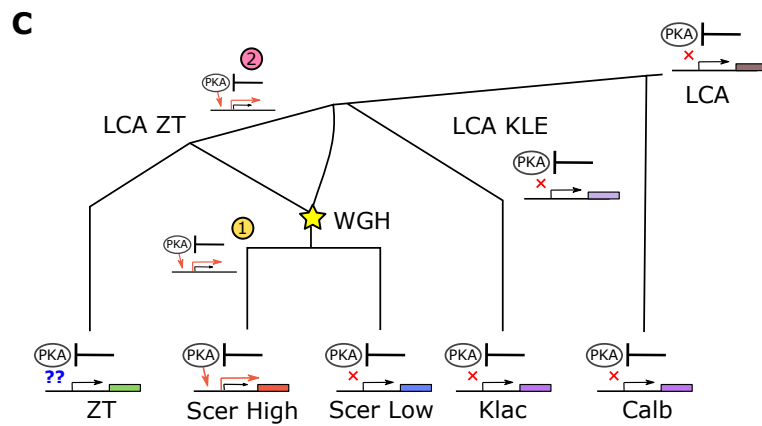
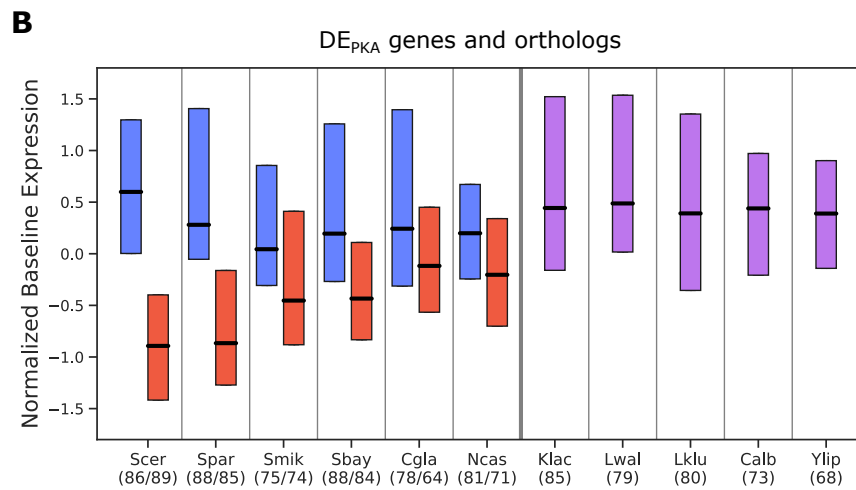
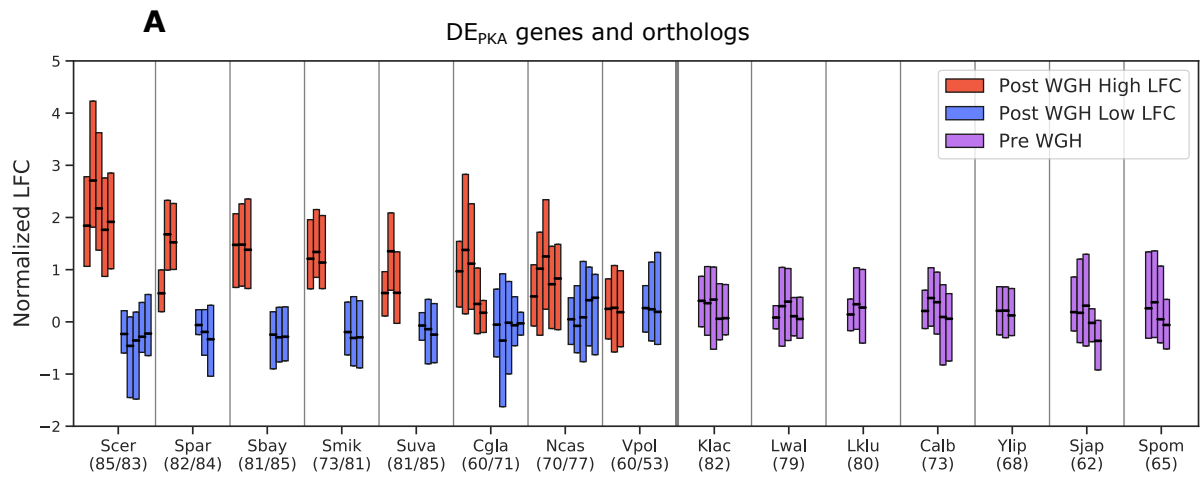


Figure 3.3: **For DE_{PKA} genes, low LFC and high basal expression is the ancestral phenotype.** (A) Boxplots showing median and Q1-Q3 range for normalized gene expression for the conditions most closely related to PKA inhibition in *S. cerevisiae* and *K. lactis* from Roy et al.⁶, and Thompson et al.⁵ (Fig 3.10). Where there are three bars, the conditions are 'DS/LOG', 'PS/LOG', and 'PLAT/LOG' from Thompson et al.⁵ and where there are five bars, the conditions are those three conditions plus 'heat shock_030' and 'heat shock_045' from Roy et al.⁶. *S. pombe* had only the three growth conditions and 'heat shock_30'. The DE_{PKA} genes in *S. cerevisiae* and their orthologs in each indicated species (when present) were analyzed for each condition. Blue and red indicate low-LFC and high-LFC ohnologs (respectively) and their syntenic orthologs in Post-WGH species. Purple bars are for the shared orthologs in Pre-WGH species. Numbers in parentheses indicate the number of retained orthologs. Syntenic ortholog assignment for Post-WGH species is based on the YGOB database¹⁸. (B) Median and Q1-Q3 range of normalized raw expression data from microarray experiments comparing mRNA under exponential growth conditions to genomic DNA from Tsankov et al.¹⁶. (C) Diagram of evolution of responsiveness to PKA for a hypothetical representative paralog pair from DE_{PKA} . The last common ancestor for all budding yeast and for the *Kluyveromyces/Lachancea/Emmenthodium* (KLE) branch is likely to have had high basal expression and little responsiveness to PKA inhibition. Responsiveness to PKA inhibition for any particular paralog pair in DE_{PKA} may have arisen either after the whole genome hybridization (labeled 1) or in the *Zygosaccharomyces/Torulaspora* (ZT) branch (labeled 2) prior to the WGH.

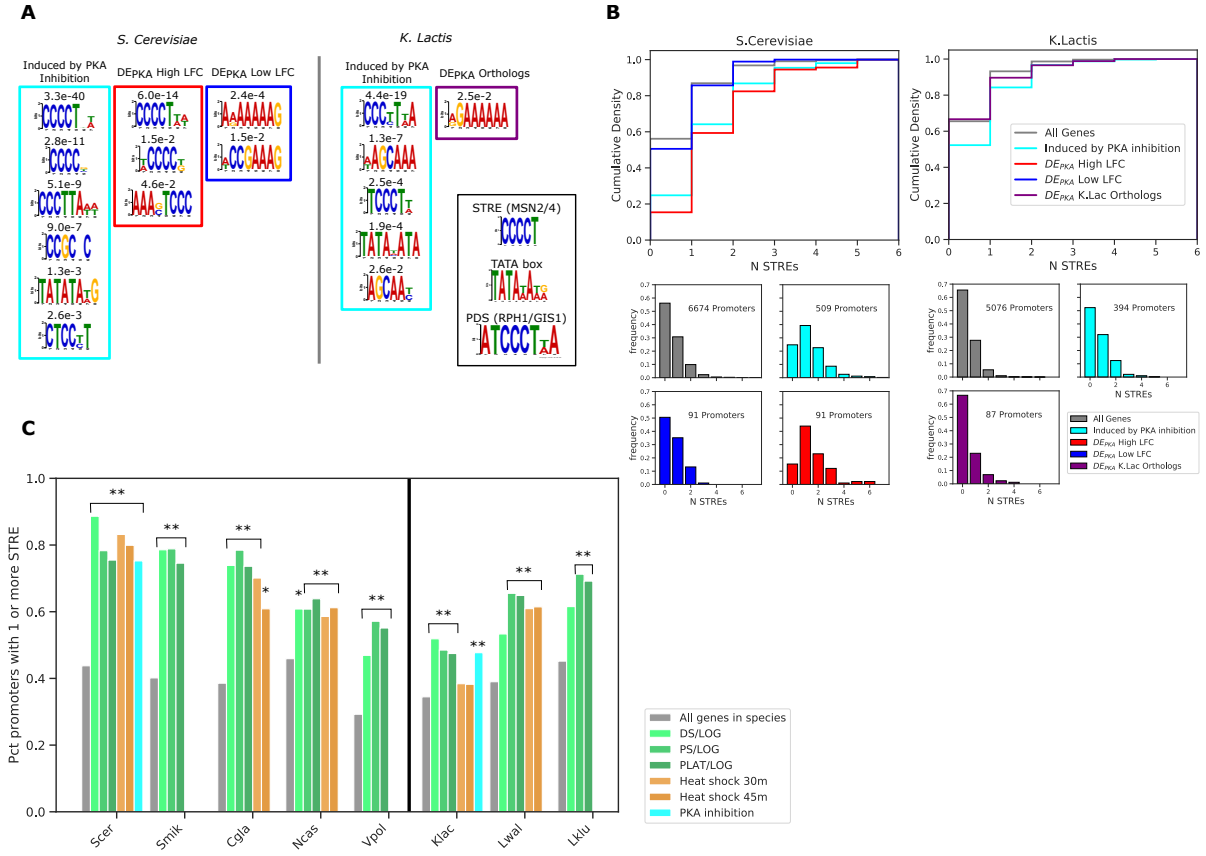


Figure 3.4: The STRE is enriched in the promoters of DE_{PKA} high LFC paralogs, and in the promoters of genes induced by stress in various species. A) Motifs identified using the DREME algorithm from the MEME suite (Bailey, 2011) in the promoters of all genes activated by PKA inhibition as well as DE_{PKA} high and low-LFC ohnolog sets (for *S. cerevisiae*) or their orthologs (for *K. lactis*) when compared against the promoters of all genes in the species. E-values are indicated above each motif. The box shows motifs for the STRE, TATA box, and PDS. B) Cumulative distributions (top) and histograms (bottom) of the numbers of STREs in the promoters of all genes activated by PKA inhibition as well as DE_{PKA} high and low-LFC ohnolog sets (for *S. cerevisiae*) or their orthologs (for *K. lactis*). C) Percentage of promoters with 1 or more STRE for the for all genes in a given species, or for genes that have LFC values above 2.5 for the indicated PKA related conditions from Fig S4. Black line separates pre-WGH from post-WGH species. Stars mark conditions for which number of STREs greater than 1 was statistically different than all genes in that species using Fisher's exact test. ** $pvalue < 1.0e^{-3}$, * $pvalue < 5.0e^{-2}$.

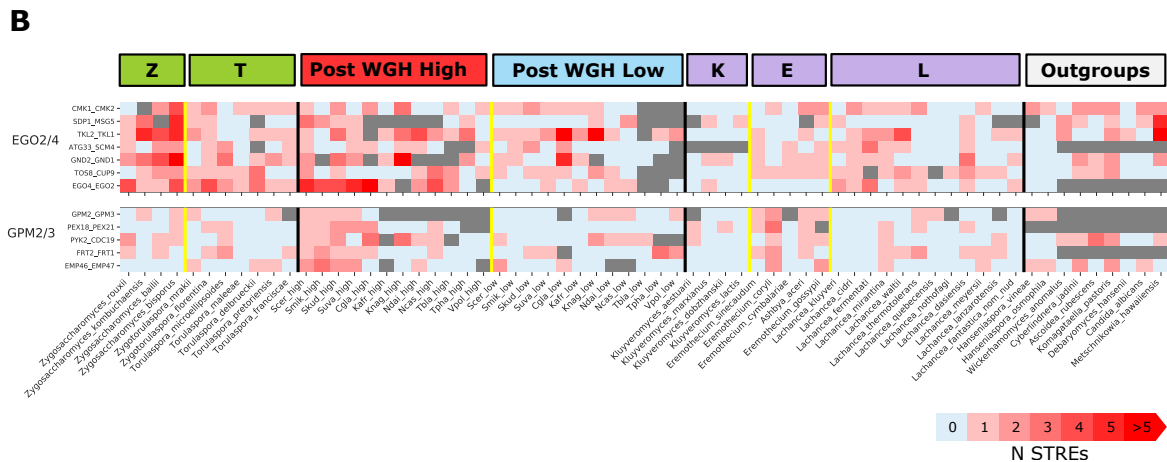
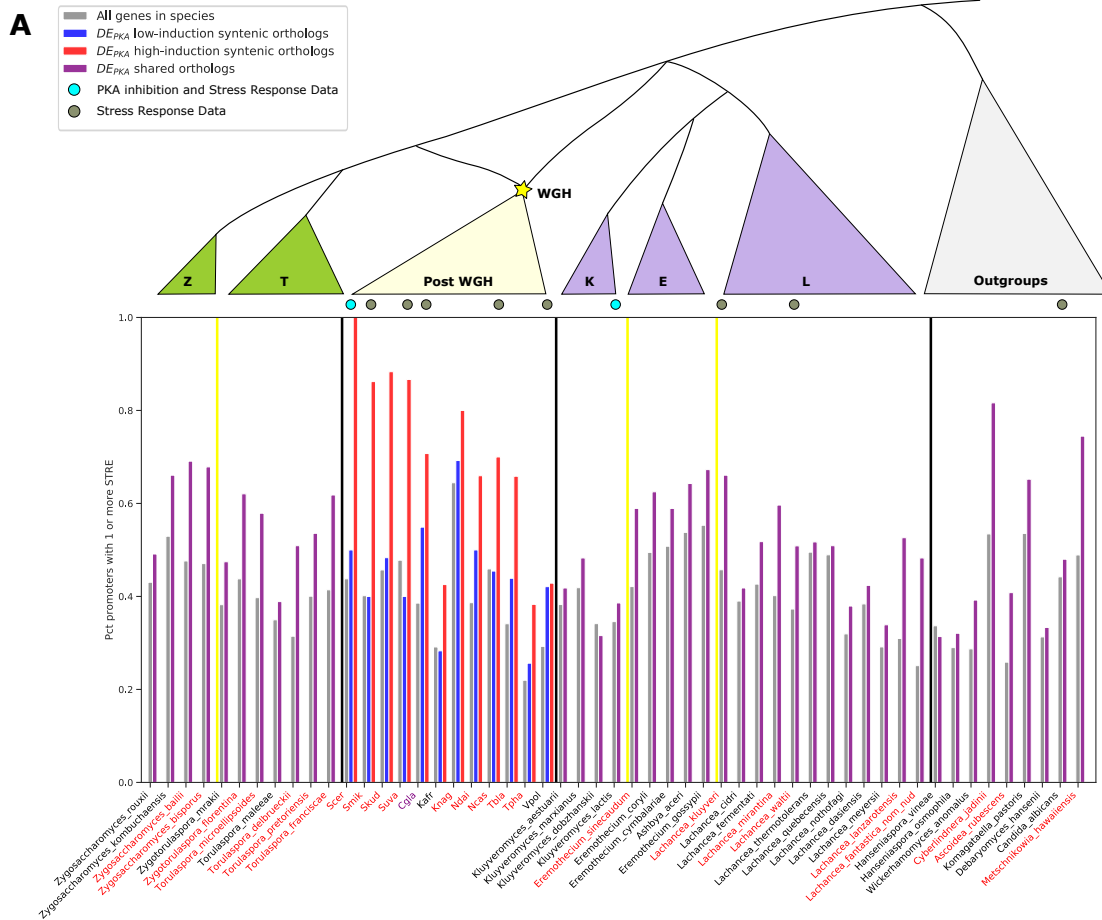


Figure 3.5: **For some DE_{PKA} high-LFC orthologs, the STRE arose in the promoter in the ZT branch prior to the WGH in some cases and following the WGH for others.** A) Percentage of promoters with 1 or more STREs in all genes in a given species (grey) or in a subset of the orthologs of DE_{PKA} genes (purple for pre-WGH species and blue and red for syntenic orthologs of the low and high LFC ohnologs respectively). For this analysis we only include ortholog pairs which had at least one ohnolog in the pre-WGH species analyzed, had no more than 8 species with duplicates in pre-WGH species, and had at least one STRE in the high-LFC ohnolog in *S. cerevisiae*. Grey dots above a species indicate that gene expression data is available for stress conditions, and light blue dots are above *S. cerevisiae* and *K. lactis* which also have data for PKA inhibition as well as for stress conditions. Red labels indicate species for which DE_{PKA} orthologs (for pre-WGH species) and syntenic orthologs of the DE_{PKA} high LFC ohnologs (for post-WGH species) were significantly enriched for promoters with 1 or more STRE relative to all genes in the species (Fig 3.20, $p < 0.05$, Fisher’s Exact Test). *C. glabrata* has a purple label indicating that both DE_{PKA} high and low LFC ohnologs were enriched for STREs. B) Number of STREs in the promoter for two clusters of ohnolog pairs in various species. Grey indicates that no ortholog was found. Data from pre-WGH species and outgroups was used as input to a hierarchical clustering algorithm (see Materials and Methods).

Figure 3.6: **Examples of the STRE arising in the promoter of the DE_{PKA} high LFC ohnolog either prior to or following the WGH.** A, B, C) Promoters of orthologs for (A) EGO2/4, (B) GND1/2 and (C) GPM2/3 for various sequenced yeast species from Shen et al.¹⁷ with STREs (red triangle) and TATA boxes (blue triangles) highlighted. The first box after each promoter represents the number of STREs and the second box indicates whether there is a TATA box within 300 bases of the start codon. Phylogeny is determined from a multiple sequence alignment of the protein sequences (see methods). Shading represents different groups of species; blue = Post-WGH, syntenic ortholog to low-LFC ohnolog ; red = Post-WGH, syntenic orthologs to high-LFC ohnolog ; yellow=Post-WGH, synteny not determined; green= ZT; light purple=KLE; dark purple=other Pre-WGH; grey=outgroups. Bar plots show rlog expression with and without 1-NM-PP1.

3.6 Materials and Methods

3.6.1 Plasmids and strain construction

All plasmids and strains used in this study are listed in Tables S[plasmids] and S[strains].

The *S. cerevisiae* TPK1/2/3(AS) base strain was a generous gift of Nan Hao and Erin O’Shea (Hao and O’Shea, 2011) and in addition to containing the gatekeeper point mutations (M164G, M147G and M165G for TPK1, TPK2, and TPK3 respectively), the strain contained an NHP6A-IRFP nuclear marker which was not used in this study.

The *K. lactis* TPK2/3(AS) strains (yBMH132, yBMH078) were constructed using a single plasmid CRISPR strategy based on Ryan and Cate⁵⁹. In order for the single Cas9/sgRNA expression vector to work in *K. lactis* we used Gibson assembly to combined Cas9 and sgRNA expression constructs on a backbone with an autonomously replicating sequence from *K. lactis* which has been shown to be expressed in a variety of budding yeast species⁶⁰. The base Cas9/sGRNA expression vector and the Pan-ARS backbone were generous gifts from Jaimie Cate and Maitreya Dunham respectively. The targeting

sequence for that guide was changed using Gibson assembly combining PCR products containing a new guide sequence with the digested backbone from a previously assembled sgRNA plasmid. Donor constructs were designed to have at least 300 bp of homology upstream and downstream from the point mutation, as well as a heterology block, synonymous mutations in the location of the sgRNA target that prevent recutting by the Cas9/sgRNA complex as described in Horwitz et al.²⁰. The donor cassette was printed by SGI-DNA, inc. and integrated into a PUCGA 1.0 backbone.

Our *K. lactis* WT strain (yLB13a, a generous gift of Alexander Johnson) was made auxotrophic for URA by counterselecting on 5-FOA. Several colonies that survived the counterselection were sequenced and were missing a 1.8 kb region of the original strain's genome that had contained the URA marker surrounded by homologous LTRs.

The transformation for the CRISPR gene editing mutation for the *K.lactis* TPK2/3(AS) strain used for RNA-seq and growth experiments (yBMH132) was performed using a standard Lithium Acetate protocol designed for transformations in *S. cerevisiae* based on the one published in Lee et al.⁸¹. Briefly, yeast colonies were grown to saturation overnight in YPD, then diluted in the morning to achieve an OD of 0.8 in 5 hours (assuming a doubling time of 110 min and lag time of 90 min). For each transformation, 4ml of cells were pelleted and washed once with water and twice with 100mM Lithium Acetate. A transformation mix was made by resuspending cells in 17 μ l water, 80 μ l 50% PEG-3350, 12 μ l 1M Lithium Acetate, and 8 μ l 10mg/ml salmon sperm DNA (Invitrogen 15632-11). 117 μ l of the transformation mix was then combined with up to 20 μ l DNA. The DNA/transformation mix was incubated at 42°C for 25-30 min, pelleted, washed with water, plated on SD-URA plates, and incubated for 2-3 days for *S. cerevisiae* and 3-4 days at 30°C for *K. lactis* strains. The DNA mix contained 5 μ g Donor DNA PCR amplified and column purified from the Donor DNA plasmid, and 1 μ g guide plasmid.

The transformation for the CRISPR gene editing mutation for the *K.lactis* TPK2/3(AS)

strain used for the KL.Msn2 nuclear localization experiment (yBMH078) was performed using a variation of a Lithium Acetate Protocol graciously provided by Jamie Cate. Cells were grown overnight to saturation and then 100ml (or 7.5ml/ transformation) of cells were grown to OD-600 of 0.8. Competent cells were made by washing twice in LATE buffer (100mM Lithium Acetate, 10mM Tris-HCL pH8.0, 0.1mM EDTA pH8.0) and resuspended in 300 μ l LATE, mixed with 300 μ l 50% glycerol, aliquoted, and frozen at -80°C. For transformation, 90 μ l thawed competent cells were mixed with 10 μ l ssDNA, 1 μ g CAS9-sgRNA plasmid, 5 μ g PCR amplified and gel purified donor DNA, 900 μ l PEG+LATE buffer (40% PEG, 100mM Lithium Acetate, 10mM Tris-HCL pH8.0, 0.1mM EDTA pH8.0) incubated for 30min at 30°C followed by 17 min heat shock at 42°C. Following heat shock, cells were washed in YPD, and recovered for 2 hours at 30°C. Plates were then washed in 1x TE buffer and plated onto SD-URA and incubated at 37°C for 12-24 hours followed by 2-3 days at 30°C. Following verification of CRISPR point mutations and deletions by sequencing, the Cas9-sgRNA plasmids / expression vectors were removed using counterselection on 5-FOA.

Plasmids for Msn2(C649S) and KL.Msn2(C623S) fluorescent reporters were constructed using restriction digestion and ligation of PCR products. The point mutations that ablate DNA binding for these transcription factors were made using quick change mutagenesis. Backbones for the *K. lactis* insertion vectors (pTS36 and pTS121) were a generous gift of Alexander Johnson. In addition to containing an mCherry based fluorescent reporter for their endogenous Msn2 transcription factors, each strain carried a Venus based fluorescent reporter for the Msn2 transcription factor from the opposite species which was not analyzed for this study.

Transformations for the *S. cerevisiae* Msn2 nuclear localization strains (yEW051 and yEW052) were done using a similar lithium acetate protocol as for yBMH132, except for the following variations. An initial amount of 2ml of cells at OD600 of 0.6 were used, the

pellet was washed and resuspended in LATE buffer and 30 μ l of cell resuspension was combined with 2 μ l salmon sperm DNA, 120 μ l 50% PEG-3350, 30 μ l LITE and 2-5 μ g digested integration plasmid in 10 μ l water. Following heat shock cells were washed with 10mM Tris-HCL pH8.0, 0.1mM EDTA pH8.0 (TE) buffer and plated on selective media.

Transformations for the *K. lactis* Msn2 Nuclear localization markers were performed using an electroporation procedure based on that described in Kooistra and Steensma⁸² with the following variations. Initially 50ml of OD 0.8 cells were used, wash and DTT buffer volumes were halved, and the volume of final resuspension in electroporation buffer was 240 μ l. For each transformation, 60 μ l resuspended cells, originating from about 12.5ml OD 0.8 cell suspension, were mixed with 5 μ l ssDNA, and 10-20 μ l DNA mix prior to electroporation, recovery, and plating. DNA mix consisted of 1.5 μ g cut and column purified integration plasmids in water.

The CRISPR deletion cassettes for *S. cerevisiae* Msn2/4 deletions were constructed using the plasmids and golden gate protocol from noa⁸³ which incorporates in vivo homologous recombination to complete the Cas9/sgRNA expression plasmid per Horwitz et al.²⁰. A similar set of plasmids was constructed to replace the backbone of the integration vector with the Pan-ARS backbone for use in *K. lactis* using golden gate cloning. Donor DNA for these constructs was constructed using a golden gate strategy to insert the donor sequence into the YTK095 backbone⁸¹. The donor sequence was designed to have 60bp homology for *S. cerevisiae* and 300bp homology for *K.lactis* to delete the Msn2/4 or KL.Msn2 proteins respectively. Because, unlike standard deletions, there would be no selection marker expressed in between the endogenous promoter and the 5'UTR, and thus nothing to prevent expressing the next open reading frame downstream of the deletion at the level of the endogenous deleted protein, we removed 250bp of the promoter of each protein targeted for deletion in addition to removing their coding sequences. Donor inserts were built using 3 sets of annealed oligos for *S. cerevisiae* or ordered as GeneBlocks (IDT) for *K. lactis*.

Transformations for the CRISPR gene deletions of Msn2/4 in *S. cerevisiae* (yBMH168, yBMH170) and KL.Msn2 in *K. lactis* (yBMH201) were performed using the same standard Lithium Acetate protocol as for yBMH132. The DNA mix for *S. cerevisiae* contained 20ng of BsmBI digested and column purified Cas9/sgRNA expression vector, 40ng of EcoRV digested and column purified sgRNA insertion vector for each mutation (Msn2 and Msn4), and 400ng PCR amplified and column purified donor DNA for each mutation. The DNA mix for *K.lactis* contained 100ng BsaI digested and column purified Cas9/sgRNA expression vector, 200ng of EcoRV digested and column purified sgRNA insertion vector, and 2 μ g PCR amplified and column purified donor DNA for each mutation.

3.6.2 Microscopy

Yeast were grown overnight to saturation in 4ml SDC. In the morning they were diluted to get OD 0.2 after 5 hours (6.5 hours for *K. lactis*) assuming a lag time of 90 min for both species and a doubling time of 115 for *S. cerevisiae* and 130 min for *K. lactis*. 96 well glass-bottom plates (Brooks Life Science Systems MGB096-1-2-LG-L) were prepared for imaging by coating with 0.25mg/ml concanavalin A (Sigma-Aldrich C2010, resuspended in distilled water) (ConA) for 30 min and washed twice with SDC prior to adding cells. Cells were sonicated gently (amplitude 1 for 3s using a Misonix S-4000 sonicator with a 1/16in microtip P/N #418) to separate clumps and incubated in appropriate wells for 30min. Following incubation, cells were washed twice with 100 μ l SDC leaving 100 μ l of SDC in the well above immobilized yeast cells. After 2-3 initial images 100 μ l of 2x perturbation media (SDC plus 1-NM-PP1 for a final concentration of 4 μ M, or a DMSO control) was added to the plate. Cells were then imaged every 2-3 minutes until the end of the experiment.

Widefield microscopy images were taken on a Nikon Ti-E inverted scope, with an incubation enclosure set to 30°C, and mercury arc-lamp illumination using RFP filters and

dichroic mirror from the Chroma 89006 ET-ECFP/EYFP/mCherry filter set. The microscope's perfect focus system was used to maintain focus on the samples throughout the experiment. The microscope was controlled with micro-manager software version 1.4.17 using the High Content Screening Site Generator Plugin⁸⁴. Images were taken with an Andor 512 pixel EMCCD camera (897 iAxone DU-897E) using a 40x objective (Nikon Planfluor 40x/0.75) and 1.5x optical zoom.

Image analysis was conducted using custom MATLAB scripts located at https://github.com/heineike02/image_analysis. Briefly, RFP images are background subtracted, smoothed, and cells are identified using the Lucy-Richardson deblurring algorithm (MATLAB function `deconvlucy`) with 5 iterations based on an image of a single typical cell of the appropriate species. For each identified cell, nuclear localization is calculated as the mean intensity of the top 5 brightest pixels to the median pixel intensity.

3.6.3 RNA sequencing

Cells were grown overnight at 30°C in YPD to saturation, and then diluted to obtain a 30ml of cells with an OD600 of 0.5 in 4-6 hours assuming a lag time of 90 min and a doubling time of 90 min for *S. cerevisiae* and 110min for *K. lactis*. The culture was then divided into two 12ml treatment and control cultures and 3.6 μ l of 10mM 1-NM-PP1 (for a final concentration of 3 μ M) or DMSO was added respectively. Each culture was incubated while shaking at 250RPM and 30°C and split into two 5ml aliquots prior to spinning down at 3850RPM (331rcf) an Eppendorf 5810R benchtop centrifuge for 3min. After spinning down, supernatants were poured out and remaining supernatant was aspirated off with a P1000 pipette. Cells were then flash frozen in liquid nitrogen and stored at -80°C.

RNA extraction was performed using the hot acid-phenol extraction protocol of Solís et al.¹⁴ with the following changes. Initial cell volume was 5ml instead of 1.5ml and thus

the initial spin for collecting the cells was done for 3min at 3850RPM (331rcf) instead of 30s at 13000RPM (15871rcf). Acid-phenol:chloroform:isoamyl alcohol (IAA) (125:24:1), pH 4.5 (AM9722) was used instead of pure Acid-Phenol because the chloroform aids in the separation of nucleic acid from proteins and lipids, and the IAA prevents foaming. The spin following heat incubation was performed at room temperature instead of at 4°C. A second 400 μ l chloroform wash was included prior to removing the aqueous phase from the phase lock tubes. 22 μ l of 3M NaOAc, pH5.2 was added instead of 30 μ l to precipitate the DNA. Ethanol Precipitation was done overnight instead of for 30 min.

Total RNA was aliquoted and stored at -80°C. Prior to library preparation, total RNA concentration was measured estimated using a Nanodrop 2000c spectrophotometer and screened for quality by electrophoresis. For electrophoresis, a formamide based loading dye was used to run samples on a 1.2% agarose gel in TBE buffer. Select samples were also checked for quality using an Agilent 2100 Bioanalyzer with an RNA 6000 Pico chip. 3' Sequencing Libraries were prepared using the Lexogen QuantSeq 3'mRNA-Seq Library Prep kit FWD using dual indices for each sample. Select libraries were checked for quality on the Bioanalyzer using a High Sensitivity DNA chip. Library concentrations were calculated using a Qbit 2.0 Fluorometer (Invitrogen), and 2.65 ng per sample were pooled and sequenced on an Illumina Hiseq 4000 sequencer to an average depth of between 245,000 and 5.6 million reads per sample (median 2.53 million). At least 3 replicates were collected for each sample and condition.

Sequencing data was trimmed, aligned, and checked for quality using the Bluebee genomics analysis pipeline. *S. cerevisiae* samples were run using the "FWD *S. cerevisiae* (R64) Lexogen QuantSeq 2.2.3" protocol, and *K. lactis* samples were run using the "FWD *K. lactis* (ASM241v1) Lexogen Quantseq2.2.3" protocol. The counts generated with these pipelines used the *saccharomyces_cerevisiae*_R64-2-2_20170117 GFF for *S. cerevisiae*, and many of the 3' sequencing reads fell into unannotated 3'UTR regions. We therefore created

a modified GFF file which included 3' UTRs from (Nagalakshmi et al., 2008). We had a similar issue for the *K. lactis* reads, but as there is no definitive studies annotating the *K. lactis* 3'UTR, we created a *K. lactis* GFF with 400bp extensions for each annotated gene serving as estimated 3'UTR regions. We then recalculated gene counts for each species using htseq in intersection-nonempty mode. Custom scripts and the updated GFF files used for these calculations are available at https://github.com/heineike02/UTR_annotation and https://github.com/heineike02/rna_seq_processing.

Gene count data was processed to yield Log Fold Change and pValue estimates using the DESeq2⁶¹ package in R (Bioconductor). Log Fold Change and pValue for PKA(AS) strains (both WT and Δ Msn2/4 or Δ KL.Msn2) in each species was calculated using all replicates in the respective strain +/- drug at 50 min. Samples which had no counts for any condition were removed. Rlog values for each replicate in *S. cerevisiae* and *K. lactis* was calculated using a DESEQ call that included data from 59 and 35 samples respectively which were sequenced in the same run. Mean rlog data for each condition was calculated by averaging rlog values across replicates. These samples included the WT +/- drug and AS +/- drug experiments as well as experiments with Msn2/4 deletion and Rph1/Gis1 deletion mutants. To calculate distributions of rlog values for all genes in *S. cerevisiae*, 443 orfs classified as dubious in the *saccharomyces_cerevisiae*_R64-2-2.20170117 GFF were filtered out. Custom scripts for further RNA sequencing analysis and enrichment are located at https://github.com/heineike02/expression_broad_data

3.6.4 GO term enrichment analysis

The GO Slim database used for analysis was downloaded from SGD on 20180412², and only Biological Process terms were analyzed. Fisher's exact test was used to test the hypothesis that the proportion of genes with a particular GO Slim term in a given set is

greater than would be expected given the distribution of that term in a background set. The background set for the genes only activated or repressed in *S. cerevisiae* was all genes in *S. cerevisiae*. The background set for the set of genes whose orthologs were also repressed in *K. lactis*, the set whose orthologs were also activated in *K. lactis*, the set which was not activated in *S. cerevisiae* but whose orthologs were activated in *K. lactis* and the set which was not repressed in *S. cerevisiae* but whose orthologs were repressed in *K. lactis* was the set of all genes in *S. cerevisiae* which contain orthologs in *S. cerevisiae*.

3.6.5 Ohnolog and ortholog assignment

Our ohnolog dataset for *S. cerevisiae* was downloaded from the Yeast Genome Order Browser (YGOB) website on 20171016. Fisher’s exact test was used to test the hypothesis that the proportion of ohnologs in a given set is greater than would be expected given the distribution of ohnologs in the background set. The same test sets and background sets were used as for GO term enrichment analysis.

Ohnolog sets for other species were identified using the “pillars” database downloaded from the Yeast Genome Order Browser on 20160908.

Orthology assignments between *S. cerevisiae* and *K. lactis* were generated using the YGOB pillars database.

We used orthology assignments from the fungal orthogroups⁴¹ for comparisons of microarray data from Roy et al.⁶ and Thompson et al.⁵ when present. That database contained no orthology assignments for *V. polymorpha*, so orthology assignments from *S. cerevisiae* to *V. polymorpha* were generated using YGOB pillars. Also the genenames used for *K. lactis* on the fungal orthogroups website and in those microarrays was inconsistent with the standard genenames which is what is used by YGOB. We therefore generated a

mapping for mismatched genenames between the two species using protein similarity based on the genome sequences provided by each study. To generate a similarity score between candidate proteins we used the `pairwise2.align.globalms` function from `biopython` with the following parameters: `match_points=1`, `mismatch_points=-1`, `gap_open=-0.5`, `gap_extension=-0.1`. We then took as candidates the top 5 genes above a particular threshold (in the case of *K. lactis*, 138). We kept the top candidate and any lower scoring candidates as long as there was not a drop in 10 points between that candidate and the next highest scoring gene.

The fungal orthogroups mappings for *N. castellii* also did not use standard genenames, so in order to generate orthologs from *N. castellii* to *V. polymorpha* to see how the orthologs of genes differentially expressed under stress in *N. castellii* behaved in *V. polymorpha* we first needed to generate a similar mapping of the *N. castellii* gene name from the fungal orthogroups index to that of YGOB. We used the same protocol as for *K. lactis* with a threshold score of 115.

To generate orthologs from *V. polymorpha* to other species with microarray data in order to analyze how orthologs of genes differentially expressed under PKA related stress conditions behaved in other species those species, we had to generate similar protein sequence based mappings from the YGOB genome for a particular species to the fungal orthogroups genome for *N. castellii*, *K. lactis*, and *S. mikatae* with threshold scores of 120, 138, and 130 respectively.

To identify syntenic orthologs of *DE_{PKA}* genes within post-WGH species in order to analyze their promoters, we used YGOB with a window of ± 8 .

To determine orthology of genes from Shen et al.¹⁷, we used orthogroups as defined by that paper's `orthomcl.clusters.txt` file.

To identify syntenic orthologs within the post-WGH species for our three example ortholog

pairs (EGO2/4, GND1/2, and GPM2/3) we used YGOB to generate a syntenic alignment with a window of +/- 8 genes around each example gene. We then extracted the surrounding genes from the sequenced genomes for all species and assigned orthology for the surrounding genes based on presence in the same orthogroup as *S. cerevisiae* and *K. lactis* genes. To do this, we also had to generate a mapping based on protein sequence from the *K. lactis* standard gene name to the gene identification numbers assigned in Shen et al.¹⁷ using the pairwise alignment procedure described above and a threshold score of 110. We then manually curated synteny based on the pattern of orthologs present.

3.6.6 Microarray data processing

Gene expression datasets from microarrays measuring the response to stress conditions in various budding yeast species from Thompson et al.⁵ (GSE36253) and Roy et al.⁶ (GSE38478) were downloaded from the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo>). The data represented Log Fold Change of intensity for the experimental condition divided by the control condition. Where there were multiple datapoints for the same gene name, the median value was used. We took the median value of technical replicates. To compare data across species, we combined all data for all conditions for a given species into a single dataset, and then used the mean and standard deviation to center and normalize all data for that species.

Microarray data estimating raw expression by comparing mRNA from cells collected during exponential growth on rich media to genomic DNA from Tsankov et al.¹⁶ (GSE22193) were downloaded from the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo>). The median of all data that had more than one spot per gene name was used for each gene name. Data for replicates were quantile normalized⁸⁵ using the implementation of

(https://github.com/ShawnLYU/Quantile_Normalize), and then the median of the replicates was used for each gene name. Before comparing data across species, the data were mean-centered and normalized by the standard deviation.

3.6.7 Identification of PKA targets

Genes were classified as activated and repressed by PKA inhibition for GO term and Ohnolog enrichment analysis in both *S. cerevisiae* and *K. lactis* as illustrated in Fig 3.8. A minimum p-value was defined $-\log_{10}(pvalue) > y_{min}$ where $y_{min} = 1.5$. An LFC threshold that depended on the p-value was chosen (so that genes with high pvalues required a larger LFC in order to be members of the set). The threshold lines are described in $(LFC, -\log_{10}(pvalue))$ coordinates by (x_1, y_1) and (x_2, y_2) for both *S. cerevisiae* and *K. lactis*. The coordinates that described the LFC threshold for genes activated by PKA inhibition were $(x_{1,act}, y_{1,act}) = (2.0, 15.0)$, $(x_{2,act}, y_{2,act}) = (2.5, 0.0)$, and those describing the LFC threshold for genes repressed by PKA inhibition were $(x_{1,rep}, y_{1,rep}) = (-2.5, 0.0)$, $(x_{2,rep}, y_{2,rep}) = (-2.0, 15.0)$ and $y_{min} = 1.5$.

3.6.8 Differential expression definition

In *S. cerevisiae* we identified differentially expressed genes by first identifying ohnolog pairs in which one member of the pair was either activated or repressed more than twofold ($LFC \geq 2.0$ for or $LFC \leq -2.0$, respectively) with a $\log_{10}(pvalue)$ less than -1.5. We then retained from this set members where the other ohnolog in the pair did not change in the same direction, specifically having LFC less than 1.5 or LFC greater than -1.5 respectively. We also further required that the difference of estimated LFC between the two paralogs be greater than 2.0 in LFC. In other species, and for DE_{Stress}^{Scer} , we identified differentially expressed genes by first defining an estimated Log Fold Change by averaging across the 3-5

(depending on the species) PKA inhibition related conditions (LFC_{est}). As in the *S. cerevisiae* RNA-seq data, we first identified ohnolog pairs in which one member of the pair was activated ($LFC_{est} > 1.8$), and retained pairs in which the other member was not activated ($LFC_{est} < 1.6$). We also required that the difference in LFC_{est} between the activated and non-activated ohnolog was greater than 1.8. These thresholds were chosen because they yielded a good overlap between $DE_{stress}^S_{cer}$ and DE_{PKA} .

3.6.9 Promoter extraction, motif enrichment, and clustering

Promoter sequences were taken to be the minimum of the [700bp] upstream of the start codon or upstream bases on the scaffold. *S. cerevisiae* promoters were taken from SGD. Promoters for *K. lactis* were extracted gene by gene using NCBI E-utilities. Promoters for YGOB species used in motif analysis for genes that were targets of PKA related stress conditions (Fig 3.10C), as well as for counting STRE motifs in post-WGH species (Fig 3.5) were extracted from the genomes for those species on the YGOB website. Promoters for genes in pre-WGH species from the genomes published in Shen et al.¹⁷ were extracted from those genomes using custom scripts available at https://github.com/heineike02/y1000plus_tools,

To analyze promoters of the orthologs of DE_{PKA} genes in 32 pre-WGH Saccharomycota species and 9 outgroups, we first identified orthologs using orthogroups calculated in Shen et al.¹⁷. We removed any ohnolog pairs that did not contain any STREs in the high-LFC ohnolog reasoning that the high-LFC ohnolog in those pairs were induced through a mechanism that doesn't require the STRE. We also removed pairs that had that had no orthologs in pre-WGH species (such as USV1/RGM1). Finally, we removed pairs that had more than one ortholog in more than 8 pre WGH species. This was to avoid ohnolog pairs that were present as duplicates before the WGH (such as the hexose transporters) and

ohnolog pairs whose pre-WGH orthologs may have undergone a duplication following the WGH, but still ancient enough to be present in several species (such as SNF3/RGT2). This left 60 of 91 DE_{PKA} ohnolog pairs.

De novo motif identification was conducted on indicated sets of promoters using the DREME algorithm from the MEME suite⁸⁶ with default promoters and a background set of all promoters from either *S. cerevisiae* or *K. lactis* as appropriate.

STRE (CCCCT) and TATA box (TATA(A/T)A(A/T)(A/G)) motifs were identified by looking for exact matches within the promoters, and enrichment was calculated using Fisher’s exact test. In order to identify evolutionary trends in STRE appearance before the WGH, we clustered DE_{PKA} ohnolog pairs based on the number of STREs in the promoters of their orthologs in 32 pre-WGH Saccharomycota species and 9 outgroups. We created a distance matrix for the 60 ohnolog pairs by using the correlation distance between two rows with missing values removed. Rows were then hierarchically clustered (scipy.cluster.hierarchy) using the ‘average/UPGMA’ linkage method. Separate clusters were defined using the fcluster method with the ‘inconsistent’ parameter set to 1.1.

3.6.10 Phylogenetic trees

Phylogenetic trees for example proteins were constructed based on the protocol from Shen et al.¹⁷. Protein sequences of all orthologs for the subject ohnolog pair were identified based on orthomcl clusters from that paper. For EGO2/4 there were two separate clusters, and so protein sequences for the two clusters were combined for the analysis. A multiple sequence alignment was created using MAAFT version 6.240⁸⁷ using `-genafpair` and `-maxiterate 1000`. These multiple sequence alignments were then trimmed with trimAL v1.4.rev22⁸⁸ with the `-gappyout` option. Phylogenetic trees were then constructed using IQ-TREE v1.5.5⁸⁹ with default parameters and outgroups specified.

3.7 Supplementary Tables

Table 3.1: **GO term enrichment for PKA targets in *S. cerevisiae* and *K. lactis***
GO term enrichment was calculated from the GO Slim Dataset downloaded from SGD on 20180412² and p-values were calculated using Fisher's exact test against a background of either all genes in *S. cerevisiae* (*S. cerevisiae* only activated and repressed groups) or all genes in *S. cerevisiae* that contain a *K. lactis* ortholog (*K. lactis* only and both *S. cerevisiae* and *K. lactis* repressed and activated groups).

	Go Terms	N subset genes with goterm	P-value	pct with go term in subset	pct with go term in background	pct of all genes with go term from subset	Example Genes	
S.Cer only activated (358 genes)	generation of precursor metabolites and energy	34	4.71E-10	9.5%	2.5%	23.6%	Cellular Respiration genes + TDH1, NDE2, GLG1/2	Activated under PKA inhibition
	response to oxidative stress	22	1.73E-06	6.1%	1.7%	22.0%	NQM1, CTT1, HSP12, ZWF1	
	carbohydrate metabolic process	30	8.56E-06	8.4%	3.3%	16.0%	NDE2, TPS2, GSY2, CAT8	
	biological process unknown	108	8.69E-06	30.2%	20.1%	9.3%	YBR085C-A, YIL100C-A, YDR193W, YOR343C, YHR193C-A	
	cellular respiration	17	2.21E-05	4.7%	1.3%	22.4%	COX5B, CYC7, CIT1, SHH4	
K.Lac only activated (187 genes)	cofactor metabolic process	22	2.05E-03	6.1%	3.0%	12.8%	SOL4, TKL2, GPD1, NDE2, ZWF1	
	biological process unknown	46	1.90E-06	24.6%	11.9%	7.8%	'FMP23', 'YLR307C-A', 'YDR248C', 'YDL186W',	
	meiotic cell cycle	20	2.19E-03	10.7%	5.2%	7.8%	RME2, UME6, SPO22, SPO16	
Activated in S.Cer and K.Lac (151 genes)	conjugation	11	5.86E-03	5.9%	2.3%	9.6%	STE18, GPA1, FUS3, STE4	
	biological process unknown	43	5.48E-08	28.5%	11.9%	7.3%	'YLR149C', 'FMP46', 'LEE1', 'UIP4', 'MHO1',	
	carbohydrate metabolic process	19	2.89E-06	12.6%	3.5%	11.0%	TSL1, SIP4, GPH1, SGA1, SIP4	
	oligosaccharide metabolic process	5	4.66E-04	3.3%	0.4%	27.8%	TPS1, TPS3, TSL1, PGM2, ATH1	
S.Cer only repressed (257 genes)	vacuole organization	9	1.47E-03	6.0%	1.7%	10.8%	ATG1/2/3/4/8/9/17, PBI2, YHR138C	Repressed under PKA inhibition
	cellular amino acid metabolic process	27	3.78E-07	10.5%	3.3%	14.3%	HIS4, ARG8, ARO7, GLN1	
	tRNA processing	19	1.02E-06	7.4%	1.8%	18.1%	TRM1/3/7/10/11/732, PUS4	
	RNA modification	17	6.36E-06	6.6%	1.7%	17.3%	TRM1/3/7/10/11/732, BUD23	
	nucleobase-containing small molecule metabolic process	20	4.11E-04	7.8%	3.2%	10.9%	URA1/5, CYC1, HXK2, TKL1	
	tRNA aminoacylation for protein translation	7	1.80E-03	2.7%	0.6%	20.0%	DED81, VAS1, SES1, THS1, ILS1	
	mitochondrial translation	58	8.46E-33	22.3%	2.5%	46.4%	MRP1/2/10/24/13/49, MRPL 1/2/3/6/7/8/9/10/11/13/15/17/20/25/28/31/32/33/35/36/38/40/44/49	
	mitochondrion organization	78	9.07E-25	30.0%	7.5%	21.0%	Overlap w/ mitochondrial translation, MTO1, TIM9/22/44, ATP23	
	RNA modification	16	1.31E-04	6.2%	2.0%	16.5%	PAP2, IME4, PUS1/6, TRM5/112	
K.Lac only repressed (260 genes)	tRNA processing	14	1.60E-03	5.4%	2.0%	13.9%	Overlap w/ tRNA processing, POP4, SEN2	
	rRNA processing	169	2.36E-87	38.4%	4.7%	71.9%	Various RPS, RPL, UTP genes	
	cytoplasmic translation	135	4.71E-74	30.7%	3.2%	84.4%	Various RPS, RPL genes, TMA19/46	
	ribosomal small subunit biogenesis	108	9.44E-58	24.5%	2.6%	83.1%	Various RPS, UTP genes, NOP6/7/14	
Repressed in S.Cer and K.Lac (440 genes)	ribosomal large subunit biogenesis	82	4.01E-42	18.6%	2.1%	78.8%	Various RPL genes, NOP2/4/7/12/15/16	
	ribosome assembly	48	2.62E-25	10.9%	1.2%	82.8%	Various RPS, RPL, RRP genes, RPF1	

Table 3.2: **Enrichment for ohnologs in PKA targets in *S. cerevisiae* and *K. lactis*** Enrichment for ohnologs in various gene sets in *S. cerevisiae* defined by activation or repression following PKA-inhibition in *S. cerevisiae* or by the gene's ortholog in *K. lactis*. Background set for *S. cerevisiae* only activated and repressed genes is all *S. cerevisiae* genes. Background set for other groups includes only *S. cerevisiae* genes with *K. lactis* orthologs.

	N subset genes that are ohnologs	P-value	pct ohnologs in subset	pct ohnologs in background	pct of all ohnologs from subset	
S. Cer activated	155	2.67E-10	30.5%	18.4%	14.7%	Activated under PKA Inhibition
S.Cer only activated	121	1.56E-11	33.8%	18.4%	11.4%	
K.Lac only activated	36	7.37E-01	19.3%	20.9%	3.5%	
Activated in S.Cer and K.Lac	34	3.49E-01	22.5%	20.9%	3.3%	
S. Cer repressed	179	4.88E-06	25.7%	18.4%	16.9%	Repressed under PKA Inhibition
S.Cer only Repressed	53	2.05E-01	20.6%	18.4%	5.0%	
K.Lac only repressed	48	8.49E-01	18.5%	20.9%	4.6%	
Repressed in S.Cer and K.Lac	126	1.55E-04	28.6%	20.9%	12.2%	
S. Cer no change	724	1.00E+00	15.9%	18.4%	68.4%	

Table 3.3: *S. cerevisiae* strains See Methods for details on strain construction. All deletion mutants remove 250bp upstream of the start codon to the stop codon.

Label	Description	Background	Plasmids used	Genotype	Transformation Protocol
HES 5-41	WT	W303 MATa ADE+, leu2, his3, trp1, ura3, can1	-	NaN	-
yBMH35 ¹¹	AS	W303 MATa ADE+, leu2, his3, trp1, ura3, can1, GAL+, psi+	-	TPK1(M164G), TPK2(M147G) and TPK3(M165G) (TPK1/2/3-AS) Nhp6a-iRFP (Kan)	-
yEW051	WT Msn2 Nuc Loc	HES 5-41	EW003, BMH049	SC.Msn2(C649S)-mCherry(TRP), KL.Msn2(C523S)-YFP (LEU), TPK1/2/3-AS, Nhp6a-iRFP (Kan)	LioAC 1
yEW052	AS Msn2 Nuc Loc	BMH35	EW003, BMH049 Cas9-sGRNA Expression Vector: pWS158	SC.Msn2(C649S)-mCherry(TRP), KL.Msn2(C523S)-YFP (LEU)	LioAC 1
yBMH168	WT Δ Msn2/4	HES 5-41	Donor DNA: Msn2: BMH137 Msn4: BMH138 Guide Plasmid: Msn2: BMH125 Msn4: BMH127	URA- Δ Msn2 Δ Msn4	LioAC 2
yBMH170	AS Δ Msn2/4	BMH35	As in yBMH168	URA- TPK1/2/3(AS) Δ Msn2 Δ Msn4	LioAC 2

Table 3.4: *K. lactis* strains See Materials and Methods for details on strain construction. All deletion mutants remove 250bp upstream of the start codon to the stop codon.

Label	Description	Background	Plasmids used	Notes	Transformation Protocol
YLB13a ⁶⁴	WT	YLB13a (MATa hm1alpha::URA3, hmra::TRP1, nej1::LEU2, trp1, leu2, metA1, uraA1)	-	-	-
yBMH128	WT (-URA)	YLB13a	-	URA-	5FOA
yBMH132	AS	YLB13a	KL.TPK2: Donor DNA: Amplified from pDON002 dCas9/sgRNA vector: ASP018. KL.TPK3: Donor DNA: Amplified from pDON006 dCas9/sgRNA vector: ASP 024	URA- KL.TPK2(M222G), KL.TPK3(M147G) (KL.TPK1/3-AS)(G3)	LioAC 3, 5-FOA
yBMH055	WT KL.Msn2 Nuc Loc	YLB13a	BMH035 BMH035, BMH059 KL.TPK2: Donor DNA: Amplified from pDON001 dCas9/sgRNA vector: ASP017. KL.TPK3: Donor DNA: Amplified from pDON006 dCas9/sgRNA vector: ASP024 Cas9-sgRNA Expression Vector: BMH120 Donor DNA: Amplified from BMH140 Guide Plasmid: BMH132	pKL.RPS25a-KL.Msn2(C623S)-mCherry HGH::TRP1 locus URA- KL.TPK1/3-AS (G2) pKL.RPS25a-KL.Msn2(C623S)-mCherry HGH::TRP1 locus pKL.RPS25a-SC.Msn2(C649S)-YFP NAT::LEU2 locus	Electroporation LioAC 2, 5-FOA, Electroporation
yBMH201	AS Δ KL.Msn2	yBMH132		URA- KL.TPK1/3-AS Δ KL.Msn2	LioAC 2

Table 3.5: **Plasmids** See Methods for details on plasmid construction.

Name	Bacterial Marker	Target Yeast Species	Backbone (Target::Marker)	Description	Addgene ID
EW003	Amp	<i>S. Cerevisiae</i>	pNH604 (TRP1::TRP)	pSC.ADH1-SC.Msn2(C649S)-mCherry	
BMH049	Amp	<i>S. Cerevisiae</i>	pNH605 (LEU2::LEU)	pSC.ADH1-KL.Msn2(C523S)-Venus	
WS158	Kan	<i>S. Cerevisiae</i>	2 μ Plasmid (URA)	Cas9-sgRNA expression vector	90517 ⁸³
WS082	Amp			sgRNA guide entry vector	90516 ⁸³
BMH125	Amp		WS082	sgRNA guide : SC.Msn2 (GCTACTGTTATTATGTGACG)	
BMH127	Amp		WS082	sgRNA guide : SC.Msn4 (AAGGTACTGCCACTACTGAG)	
BMH137	Amp		YTK095	Δ SC.Msn2 Donor	
BMH138	Amp		YTK095	Δ SC.Msn4 Donor	
ASP017	Amp	<i>K. Lactis</i> budding yeast	Pan-ARS plasmid (URA)	Cas9-sgRNA vector: KL.Tpk2 Guide A (CTGCATAAACTTAGCCACA) KL.Tpk2(M222G) Donor A	
DON001	Amp		pUCGA 1.0	Heterology Block: PVAKFYA(160-166) CCT GTG GCT AAG TTT TAT GCA \rightarrow CCA GTT GCC AAA TTC TAC GCT	
ASP018	Amp	<i>K. Lactis</i> budding yeast	Pan-ARS plasmid (URA)	Cas9-sgRNA vector: KL.Tpk2 Guide B (TGAAAGACGGATGTTAAAGC) KL.Tpk2(M222G) Donor B	
DON002	Amp		pUCGA 1.0	Heterology Block: DERRMLKL(109-116) GAT GAA AGA CGG ATG TTA AAG CTG \rightarrow GAC GAG AGG AGA ATG CTA AAA TTG	
ASP024	Amp	<i>K. Lactis</i> budding yeast	Pan-ARS plasmid (URA)	Cas9-sgRNA vector: KL.Tpk3 (CCTTTCATAATTAGAATGTG) KL.Tpk3(M147G) Donor Plasmid	
DON006	Amp		pUCGA 1.0	Heterology Block: PFHIR(203-207) CCT TTC ATA ATT AGA \rightarrow CCA TTT ATC ATA AGG	
BMH035	Amp	<i>K.Lactis</i>	TS36 TRP1::HGH	pKL.RPS25a-KL.Msn2(C623S)-mCherry	
BMH059	Amp	<i>K.lactis</i>	TS121 LEU2::NAT	pKL.RPS25a-SC.Msn2(C649S)-Venus	
BMH120	Kan	<i>K. Lactis</i> budding yeast	Pan-ARS(URA) / WS158	Cas9-sgRNA expression vector	
BMH119	Amp		WS082	sgRNA guide entry vector	
BMH132	Amp		BMH119	sgRNA guide : KL.Msn2 (AATCGACGGAATACCATACG)	
BMH140	Amp		YTK095	Δ KL.Msn2 Donor	

3.8 Supplementary Figures

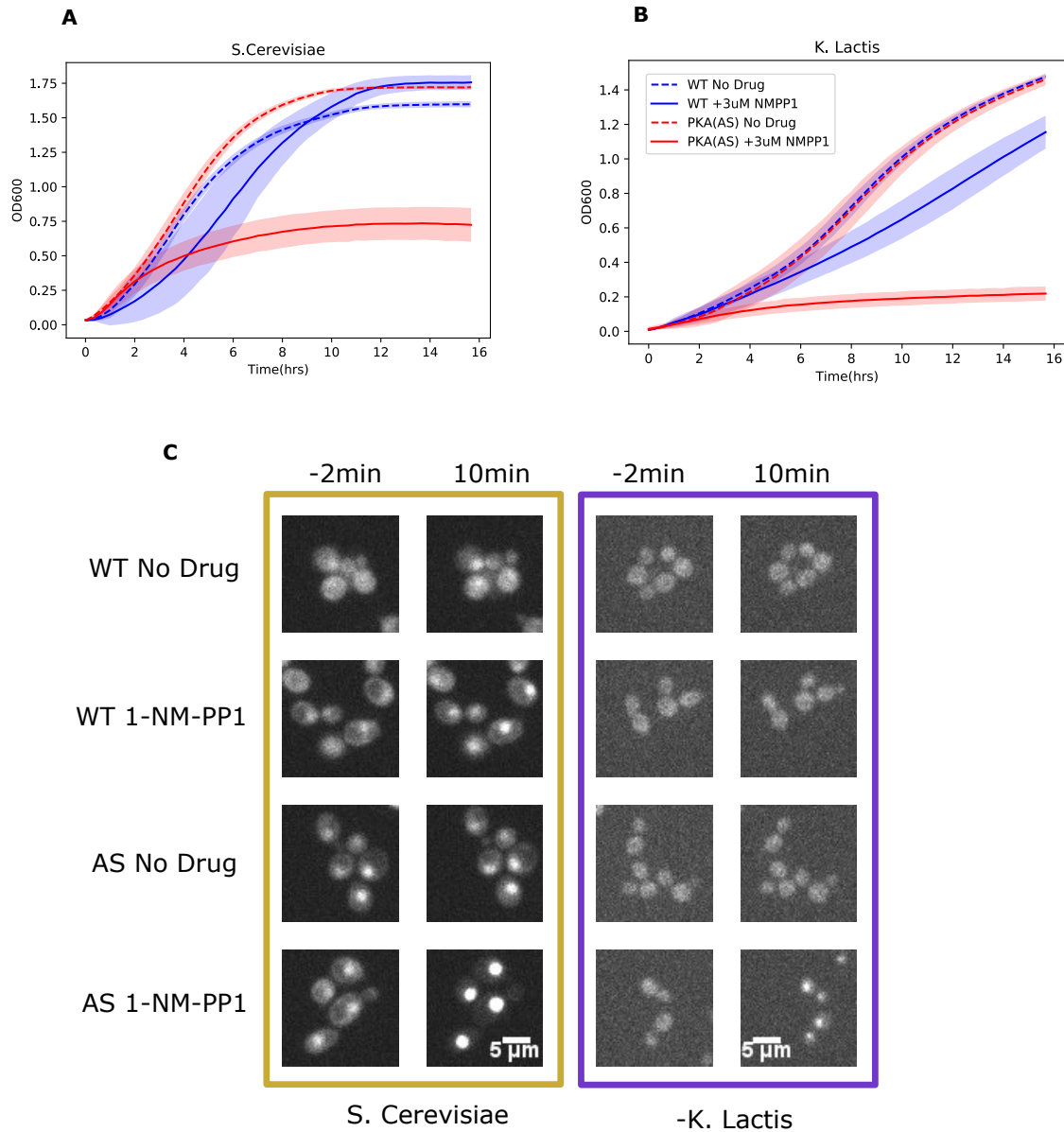


Figure 3.7: **PKA inhibition inhibits growth and causes Msn2 nuclear localization in *S. cerevisiae* and *K. lactis*.** A) WT and PKA(AS) *S. cerevisiae* and *K. lactis* strains were grown in YPD in the presence or absence of 3uM 1-NM-PP1 and OD600 was measured every 20min. Standard Deviation of at least 4 technical replicates is shown. C) Microscope images from data in Fig1 A) and B). WT and AS strains were grown in SDC and imaged either 2 minutes before or 10 minutes after adding control media or 4 μ M 1-NM-PP1.

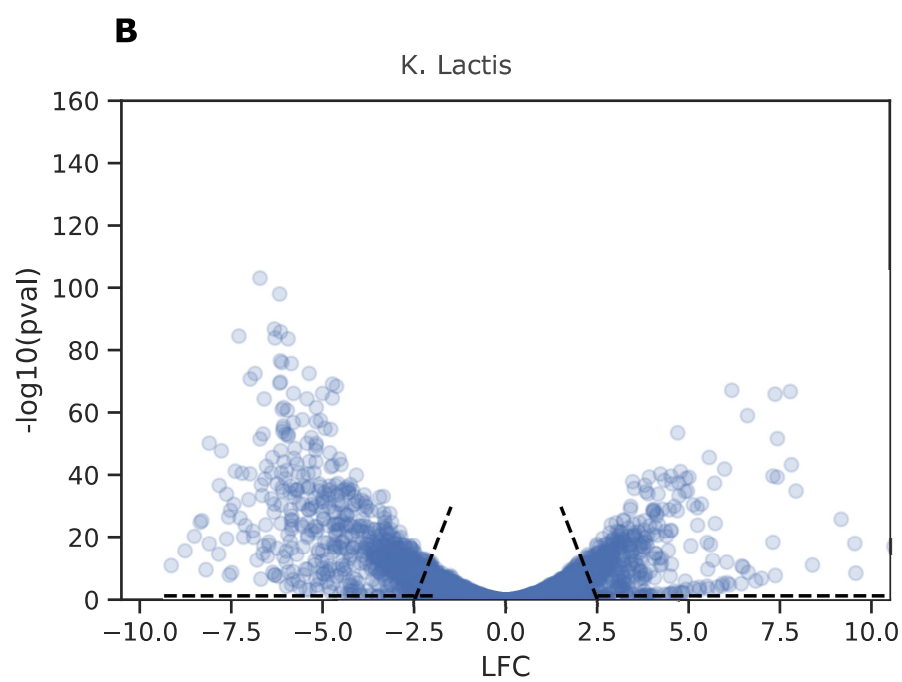
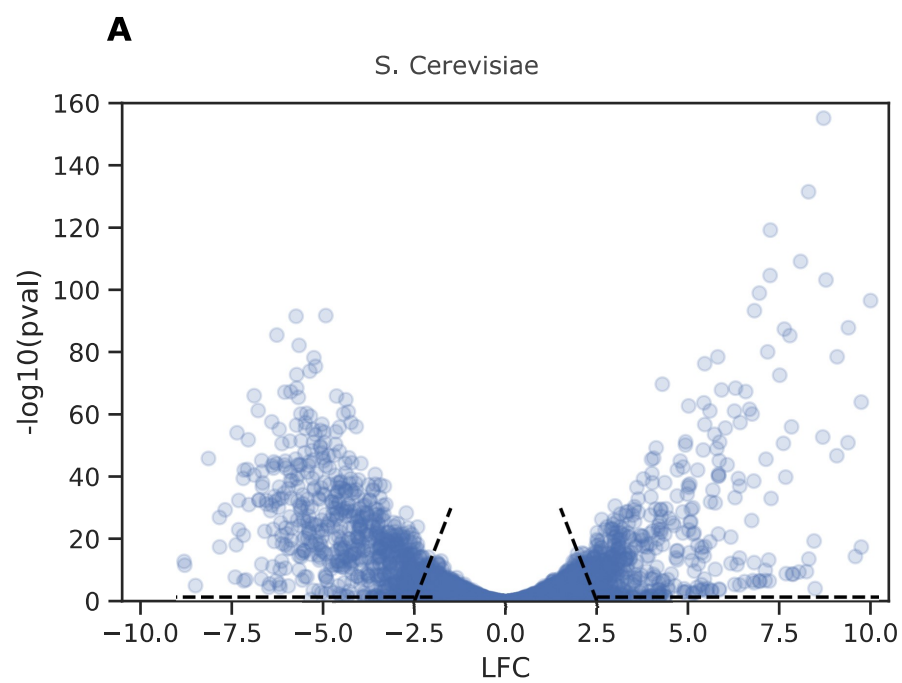


Figure 3.8: **Thresholds defining genes activated and repressed by PKA inhibition in *S. cerevisiae* and *K. lactis*.** LFC v.s. $-\log_{10}(pvalue)$ for targets of activation and repression in each species were defined as all genes that fell above the line defined by the points (x_1, y_1) and (x_2, y_2) , and for which $-\log_{10}(pvalue) > y_{min}$. For both A) *S. cerevisiae* and B) *K. lactis*, $(x_{1,act}, y_{1,act}) = (2.0, 15.0)$, $(x_{2,act}, y_{2,act}) = (2.5, 0.0)$, $(x_{1,rep}, y_{1,rep}) = (-2.5, 0.0)$, $(x_{2,rep}, y_{2,rep}) = (-2.0, 15.0)$ and $y_{min} = 1.5$.

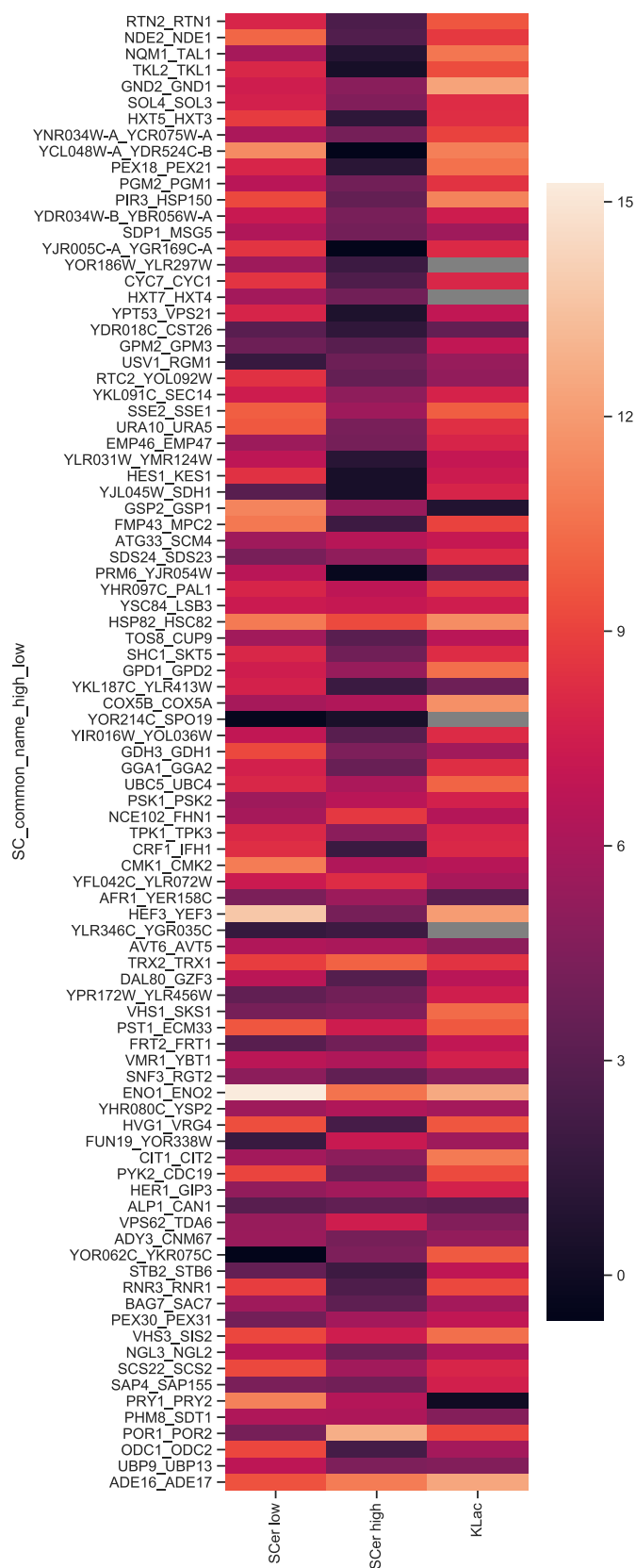


Figure 3.9: **Basal expression (rlog) for DE_{PKA} low and high LFC ohnologs and their shared orthologs in *K. lactis*.** Data are from PKA(AS) strains with no drug during exponential growth. Gray boxes indicate missing orthologs in *K. lactis*.

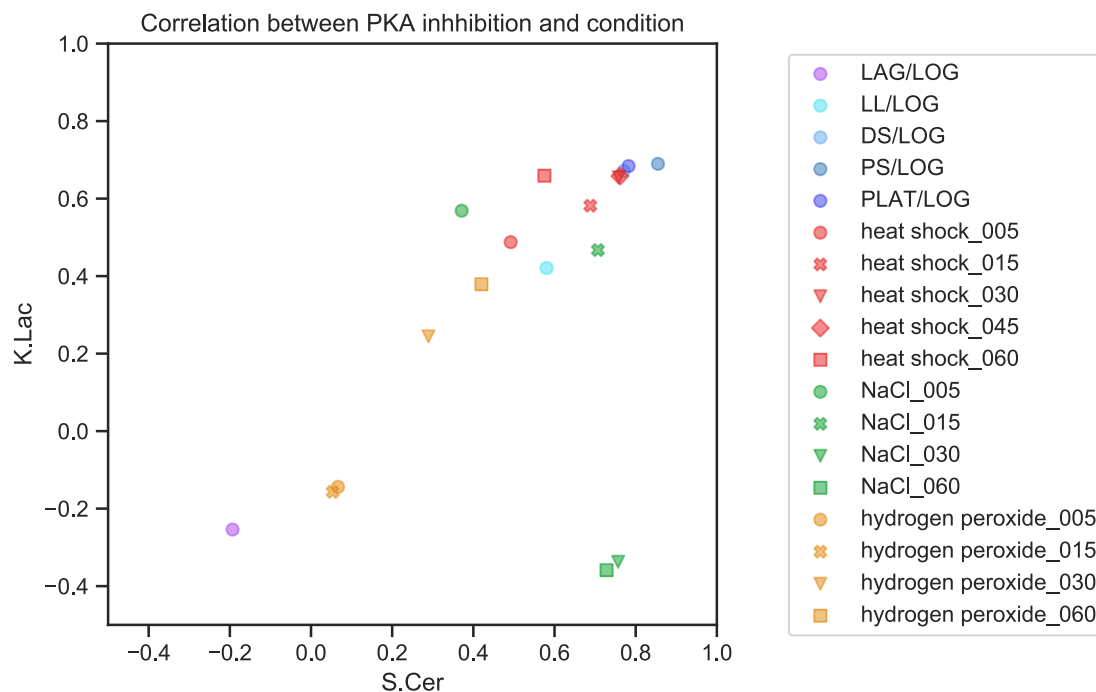


Figure 3.10: **Correlation between PKA inhibition and stress gene expression data** The stress conditions Diauxic Shift (DS), Post Diauxic Shift (PS), Plateau (PLAT), and heat shock are most closely correlated with PKA inhibition in both *K. lactis* and *S. cerevisiae*. Pearson's correlation coefficient between LFC for the indicated condition in Roy et al.⁶ and Thompson et al.⁵ and our PKA inhibition data is plotted for *S. cerevisiae* (x-axis) and *K. lactis* (y-axis).

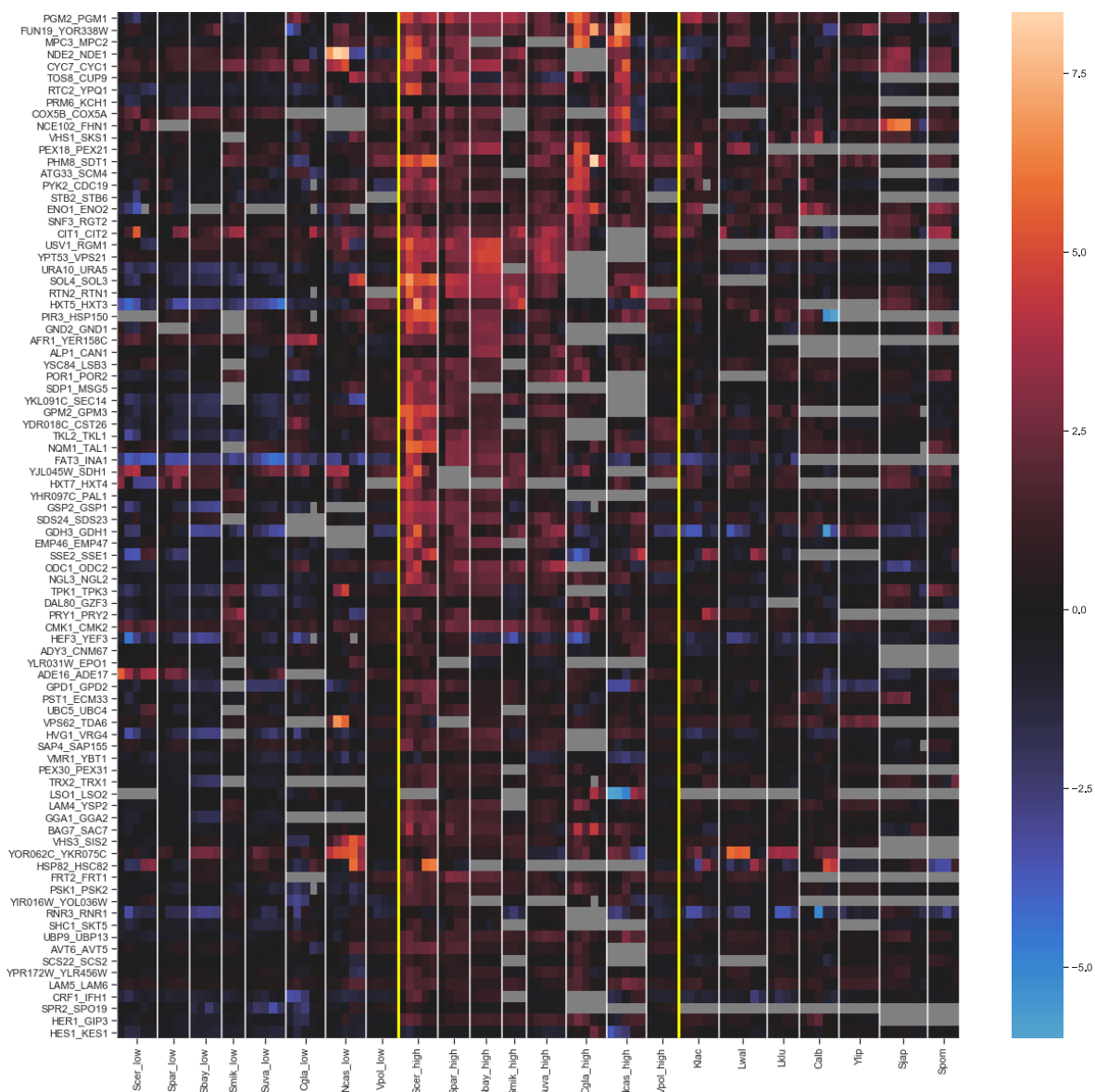


Figure 3.11: **Induction of DE_{PKA} genes during growth to saturation and in response to stresses related to PKA inhibition** Yellow lines separate pre-WGH species (on the right) and post-WGH species (two groups on the left). Syntenic orthologs of low-LFC ohnologs are on the left, and syntenic orthologs of high-LFC ohnologs are in the center. Where there are three columns, the conditions are 'DS/LOG', 'PS/LOG', and 'PLAT/LOG' from Thompson et al.⁵ and where there are five bars, the conditions are those three conditions plus 'heat shock_030' and 'heat shock_045' from Roy et al.⁶. *S. pombe* had the three growth conditions and 'heat shock_30'. The DE_{PKA} genes are arranged such that their estimated response to PKA inhibition in the high-LFC ohnolog is more conserved towards the top and less conserved towards the bottom. Estimated response to PKA inhibition was taken to be the average of gene induction over all 3-5 conditions. The ohnolog pairs (YDR034W-B, YBR056W-A), (YCL048W-A, YDR524C-B), (CIS1, YGR035C), (EGO4, EGO2), and (YOR186W, YLR297W) in DE_{PKA} did not have sufficient data in the microarray experiments to be included in the analysis.

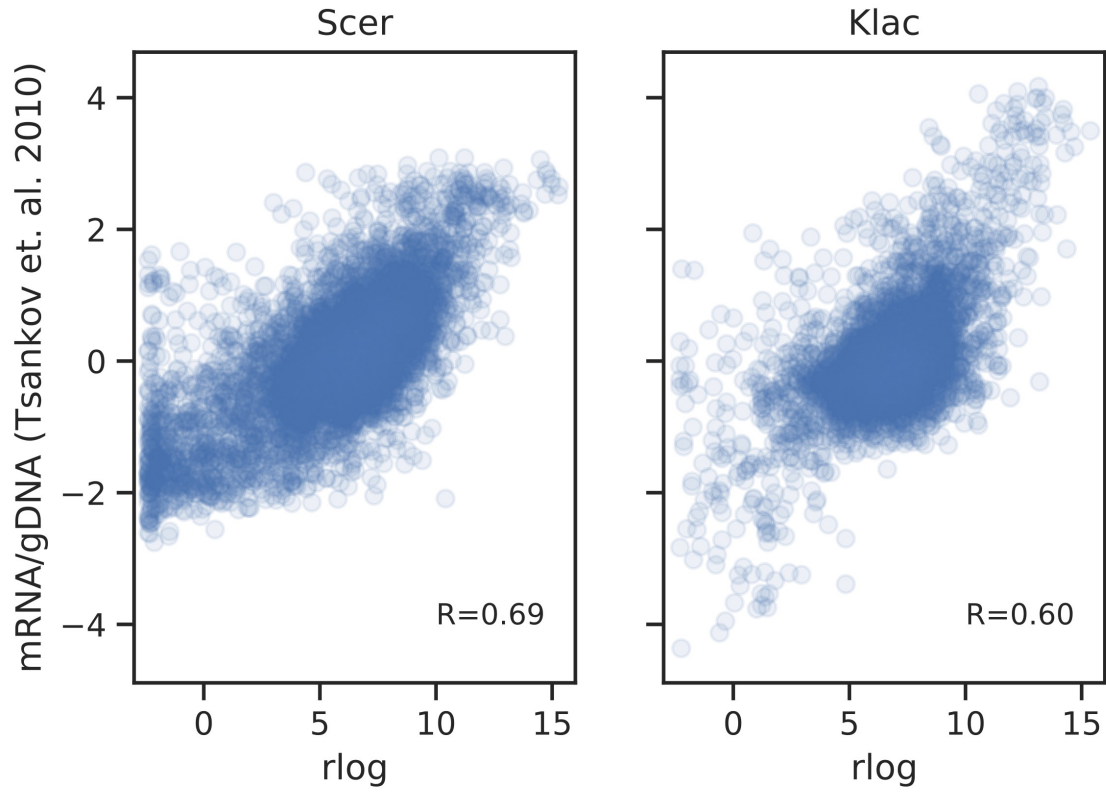
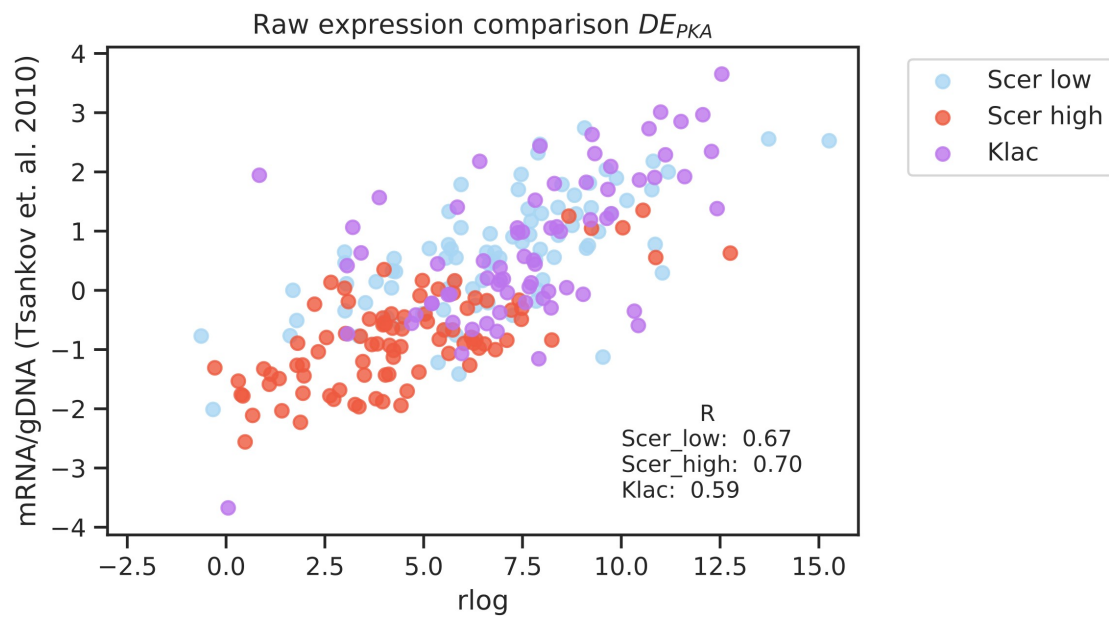
A**B**

Figure 3.12: Comparison of RNA-seq and microarray basal expression data The basal expression data from our RNA-seq experiment are correlated with basal expression data from Tsankov et al.¹⁶. A) rlog data from a PKA(AS) under exponential growth in YPD with no drug is shown on the x-axis, and basal expression from Tsankov et al.¹⁶ is shown on the y-axis for all genes in *S. cerevisiae* (left panel) and *K. lactis* (right panel). B) The same data as in A for Low-LFC (blue) and High-LFC (red) DE_{PKA} ohnologs from *S. cerevisiae* and their shared *K. lactis* ortholog (purple) are shown on the same plot. Raw expression data was mean subtracted and normalized by standard deviation (See Materials and Methods).

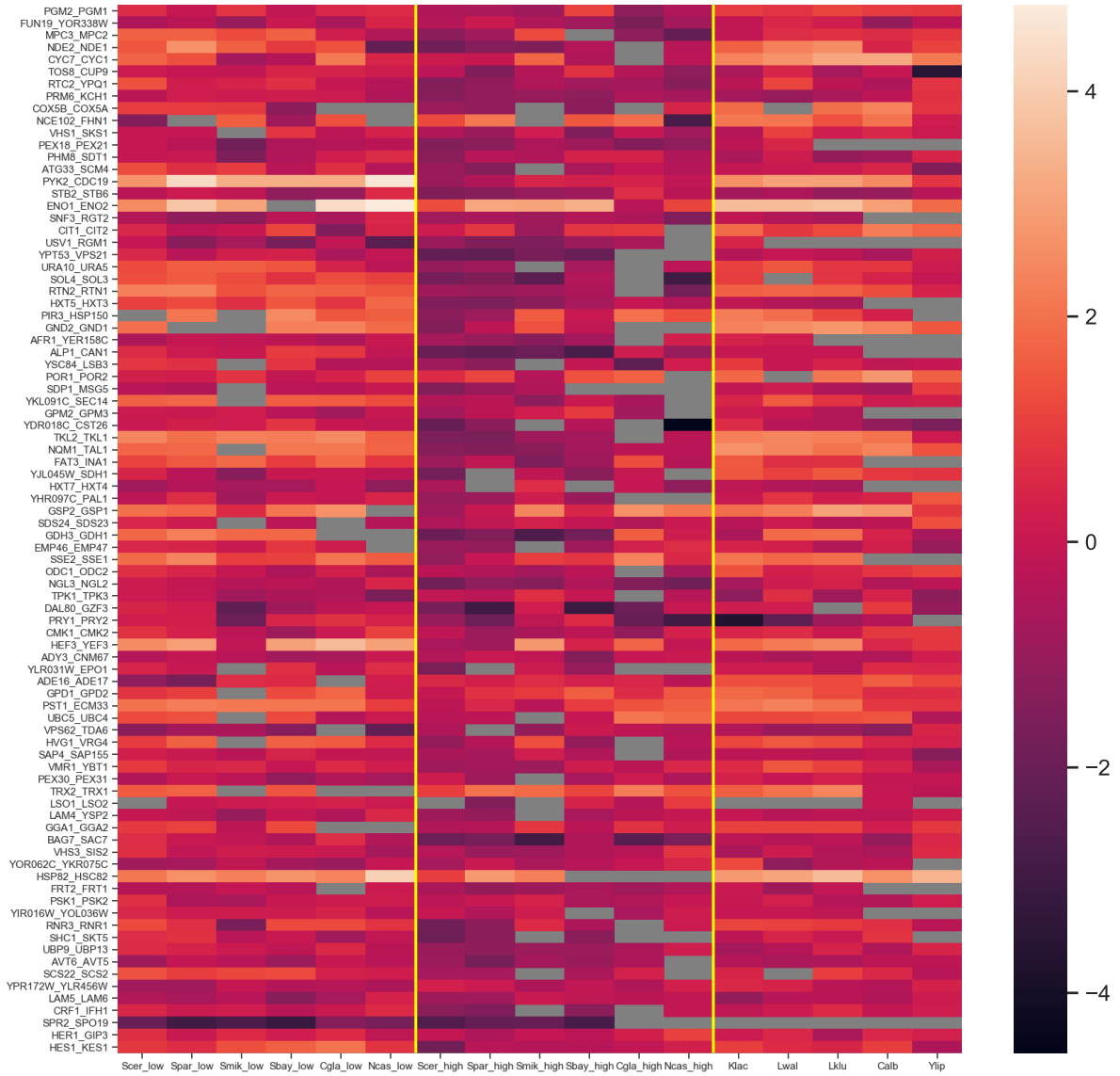


Figure 3.13: **Basal expression across DE_{PKA} orthologs for 11 budding yeast species.** Rows and columns are ordered as in Fig 3.11. Basal expression data is from Tsankov et al.¹⁶. Data from each species was mean subtracted and normalized by standard deviation (See Materials and Methods).

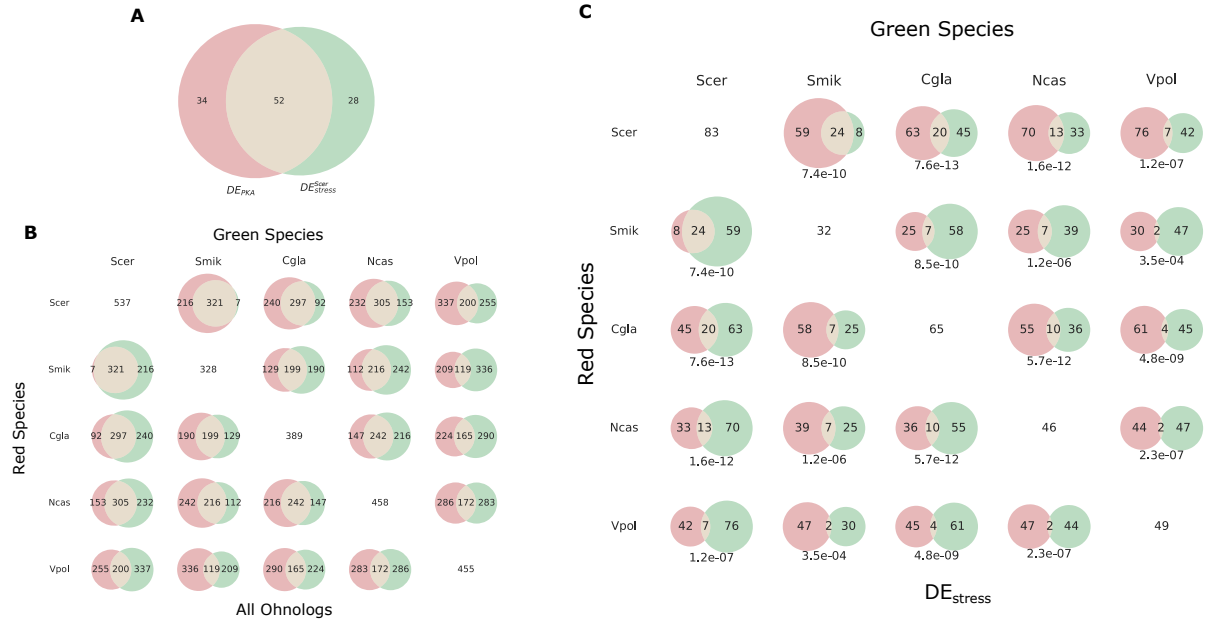


Figure 3.14: Different sets of genes are differentially expressed in different post-WGH species. We can define ohnolog pairs that have one member activated by PKA-related stress conditions (Fig 3.10) and which have differential expression in a similar manner as we defined DE_{PKA} , which we denote for a particular species as $DE_{Stress}^{Species}$ (See Materials and Methods). There is little overlap between DE_{Stress} sets defined for different species. This is partially a result of the fact that the set of ohnolog pairs that is retained decreases as the evolutionary distance between species increases for post-WGH species⁹⁰. However, the overlap of differentially expressed ohnolog pairs is even smaller than would be expected given the total percentage of retained ohnologs between species. Thus there is no reason to expect a priori that the same conservation patterns for LFC and basal expression would hold for differentially expressed ohnolog pairs defined by their expression in distantly related post-WGH species (e.g. DE_{Stress}^{Scer} v.s. DE_{Stress}^{Vpol}). A) Overlap between DE_{PKA} and DE_{Stress}^{Scer} . DE_{Stress} genes were defined based on the average normalized LFC across the 3-5 PKA inhibition related conditions (LFC_{est}) as ohnolog pairs for which one ohnolog was activated ($LFC_{est} > 1.8$), and the other was not ($LFC_{est} < 1.6$). We also required that the difference in LFCest between the activated and non-activated ohnolog was greater than 1.8. The overlap between the two sets (52 shared ortholog pairs) is statistically significant ($p=1.29E-27$, Fisher's exact test) B) Overlap of all ohnologs that have data in Roy et al.⁶ and Thompson et al.⁵ between species as determined from YGOB pillars¹⁸. The number on the diagonal is the total number of ohnologs with data in that species. C) Overlap between DE_{Stress} sets for various pairs of species. The p-value for a Fisher's exact test on the null hypothesis that this overlap is expected based on the overlap of all ohnologs with data (from B) for those two species is shown below each Venn diagram. The number on the diagonal is the total number of ohnolog pairs in DE_{Stress} for that species.

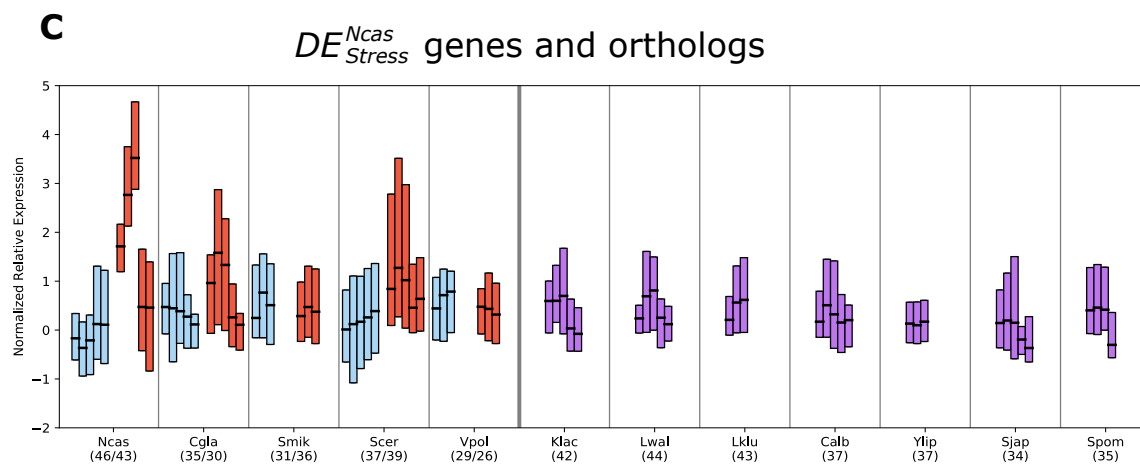
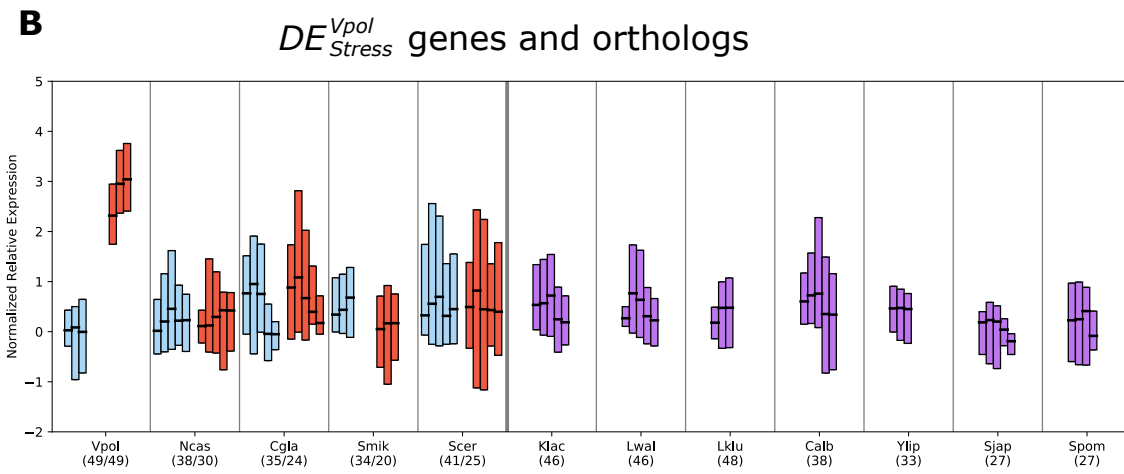
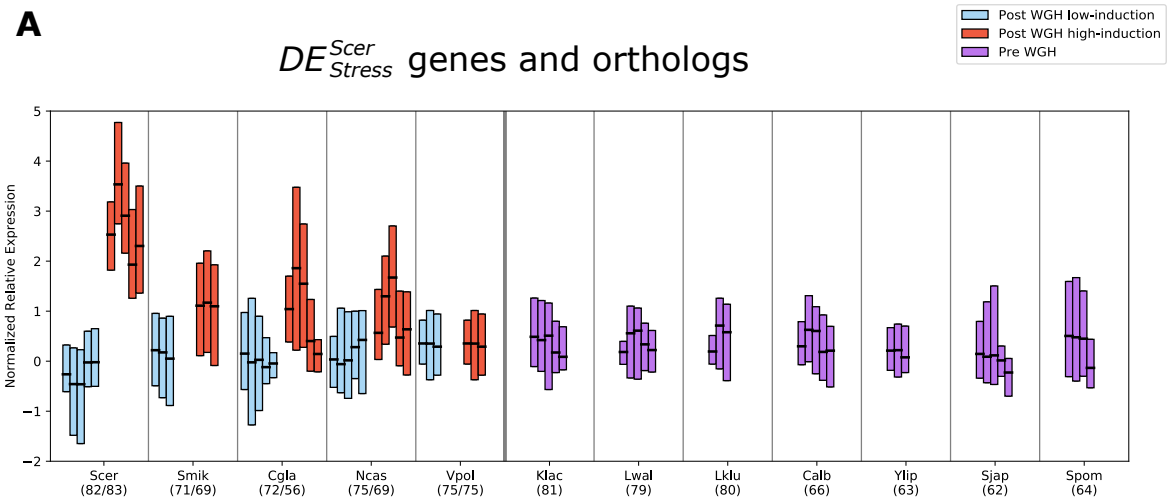


Figure 3.15: **The conclusion that PKA induction is the derived phenotype is independent of the species in which DE_{Stress} genes are defined.** Distribution of LFC values for PKA related stress conditions as in Fig 3.3A is shown, except focusing on A) DE_{Stress}^{Scer} , B) DE_{Stress}^{Vpol} , and C) DE_{Stress}^{Ncas} .

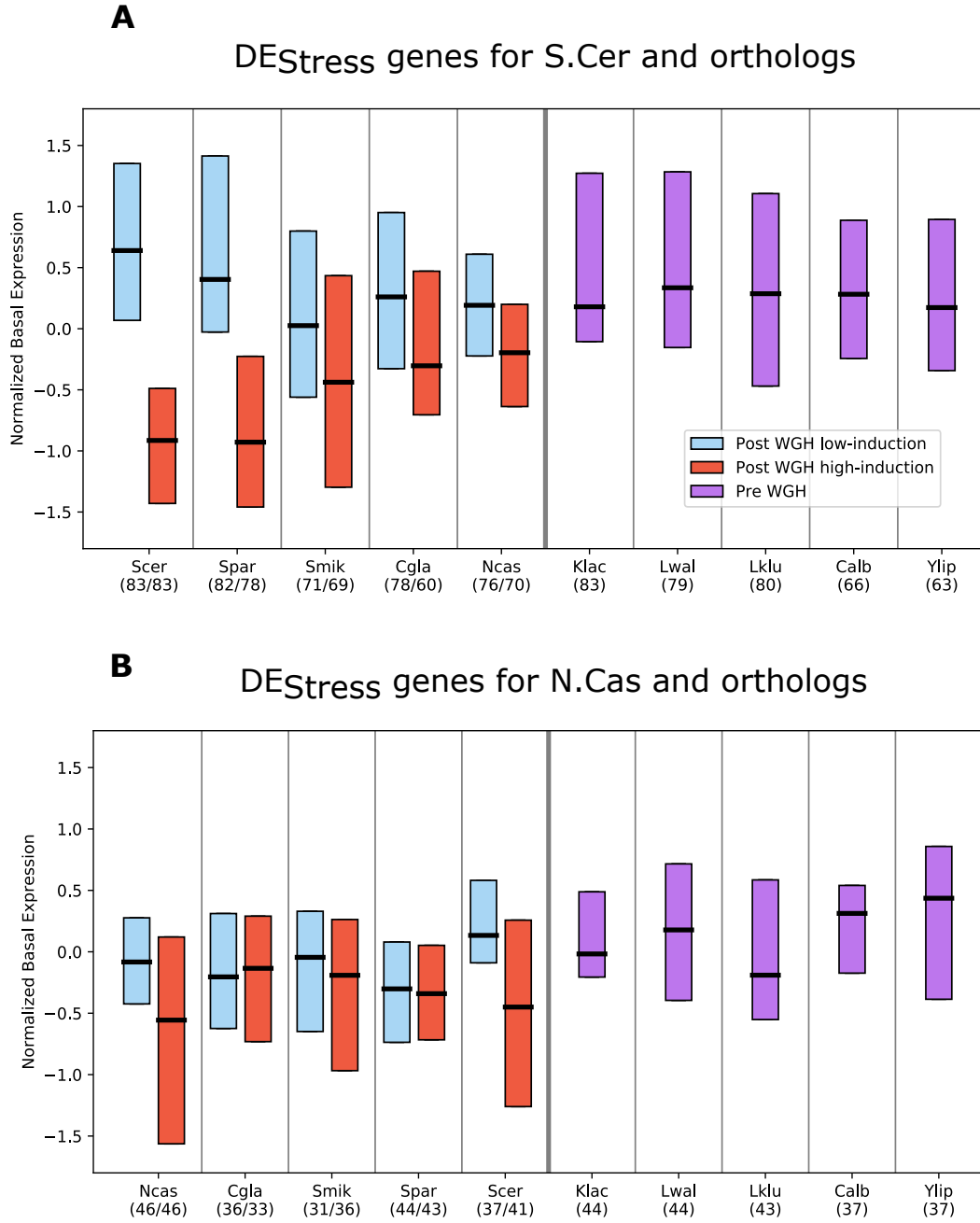


Figure 3.16: The conclusion that high basal expression is the ancestral phenotype is independent of the species in which the DE_{Stress} genes are defined. Distribution of basal expression values as in Fig 3.3B is shown, except focusing on A) DE_{Stress}^{Scer} and B) DE_{Stress}^{Ncas} .

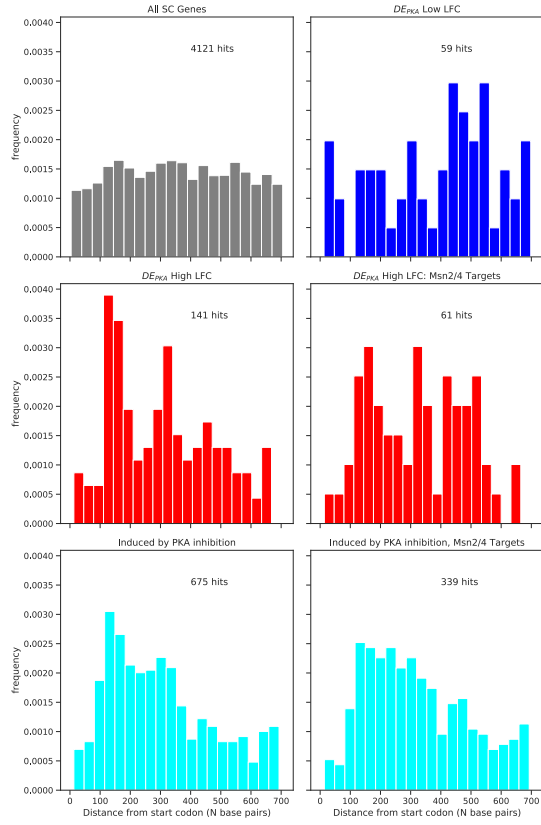
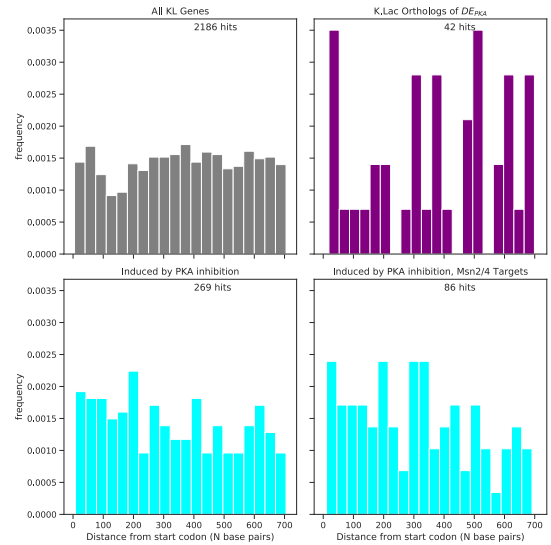
A**B**

Figure 3.17: **Location distribution of STREs in promoters** The STRE is localized closer to the start codon in the promoters of targets of PKA inhibition in *S.cerevisiae*. Distribution of STRE distance from start codon for indicated gene sets for A) *S. cerevisiae* and B) *K.lactis*.

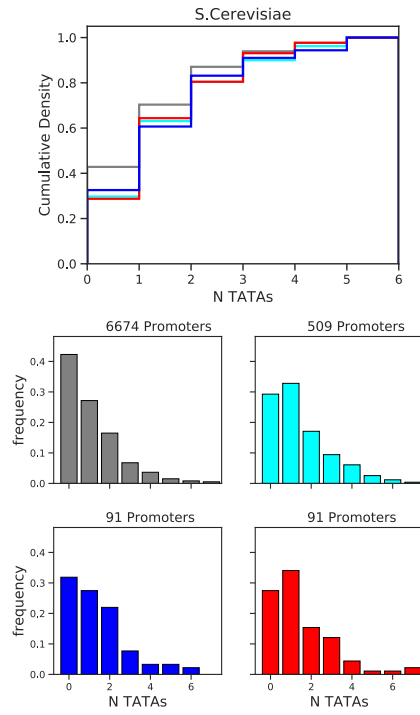
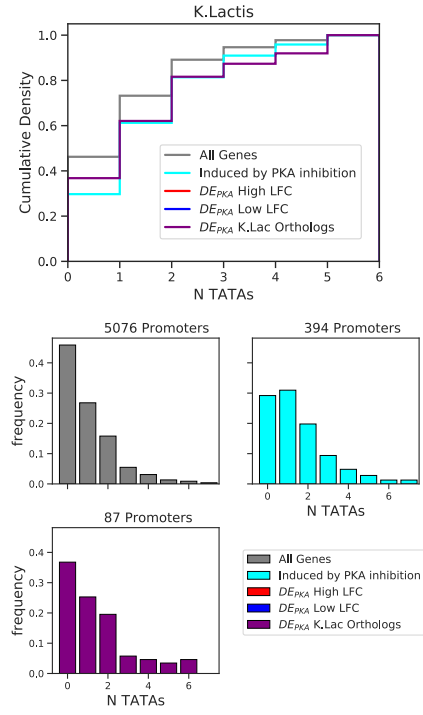
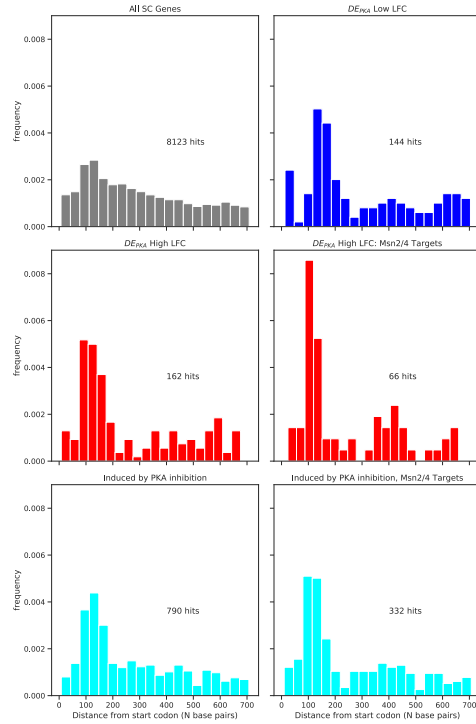
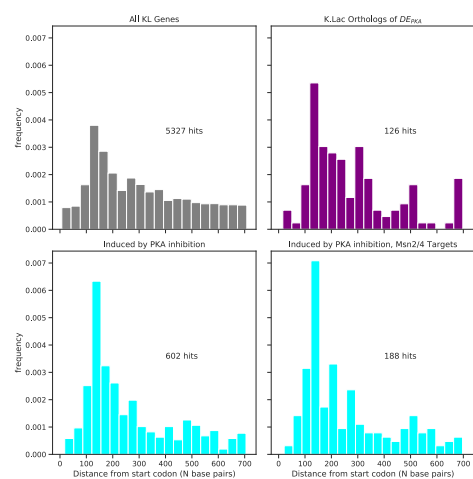
A**B****C****D**

Figure 3.18: **Distribution of number and location of TATA boxes in promoters** The TATA box is enriched for targets of PKA inhibition, as well as in orthologs of DE_{PKA} genes in *S.cerevisiae* and *K.lactis*. A) Distribution of number of TATA boxes in the promoters of indicated sets in *S.cerevisiae*. B) same for *K. lactis*. C) Distribution of the distance of TATA boxes from the start codon in the promoters of indicated sets for *S. cerevisiae* and D) *K. lactis*.

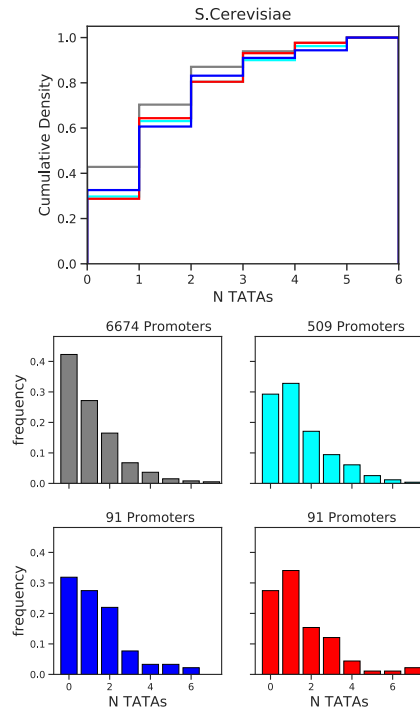
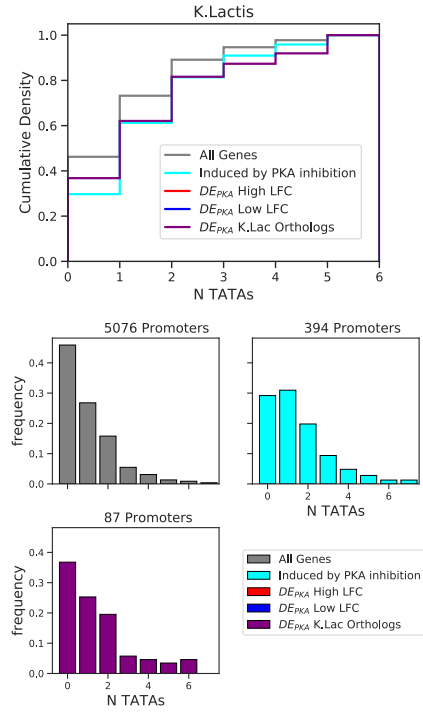
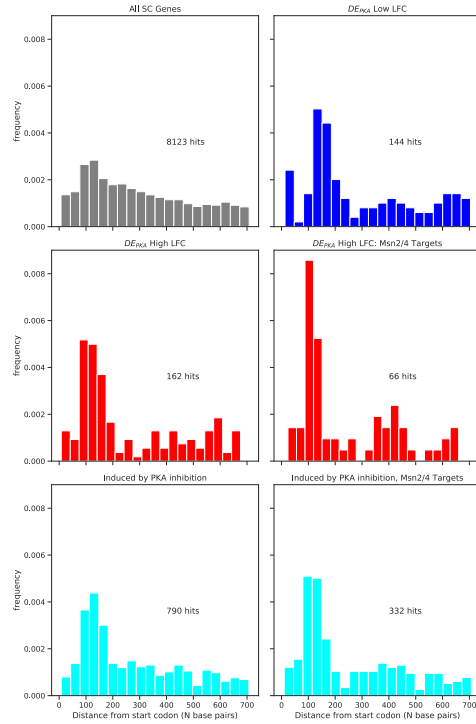
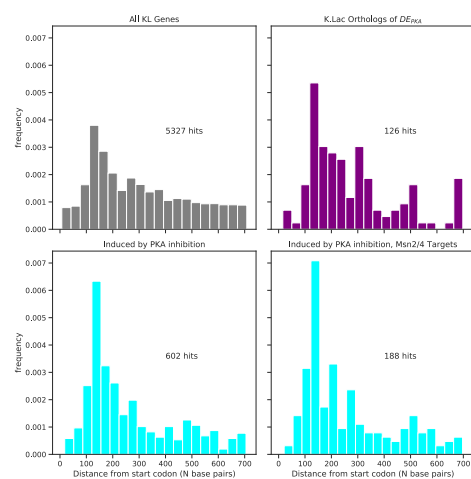
A**B****C****D**

Figure 3.19: **Joint distribution of STREs and TATA boxes** There is a high percentage of promoters with both TATA boxes and STREs in genes induced by PKA in *S. cerevisiae*, which is expected based on the enrichment for both motifs in that set. A) Percentages of promoters in the indicated sets with one or more STRE in combination with one or more TATA box in the 300bases upstream of the start codon for *S.cerevisiae* and B) *K. lactis*.

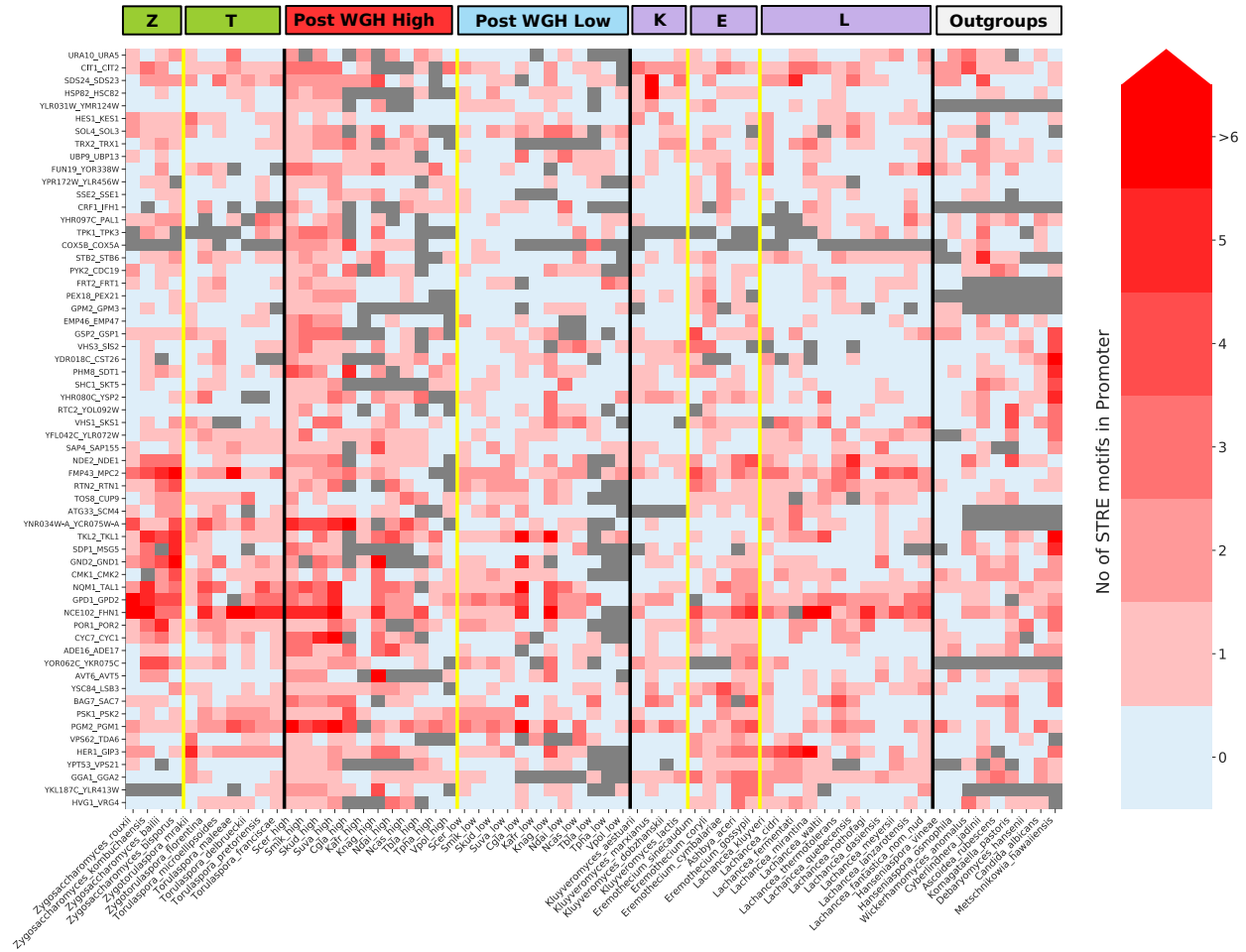


Figure 3.21: **STRE counts for promoters of DE_{PKA} orthologs** There is enrichment for STREs in the promoters of the ZT orthologs of select DE_{PKA} genes, and in the syntenic orthologs of the DE_{PKA} high-LFC ohnologs. The number of STREs in the promoters of the orthologs of the subset of DE_{PKA} genes considered for Fig S14 are shown. Columns are different species and rows are clustered based on the number of STREs in pre-WGH species (ZT branch, KLE branch, and outgroups). Grey indicates either no ortholog exists or no promoter was found in the dataset.

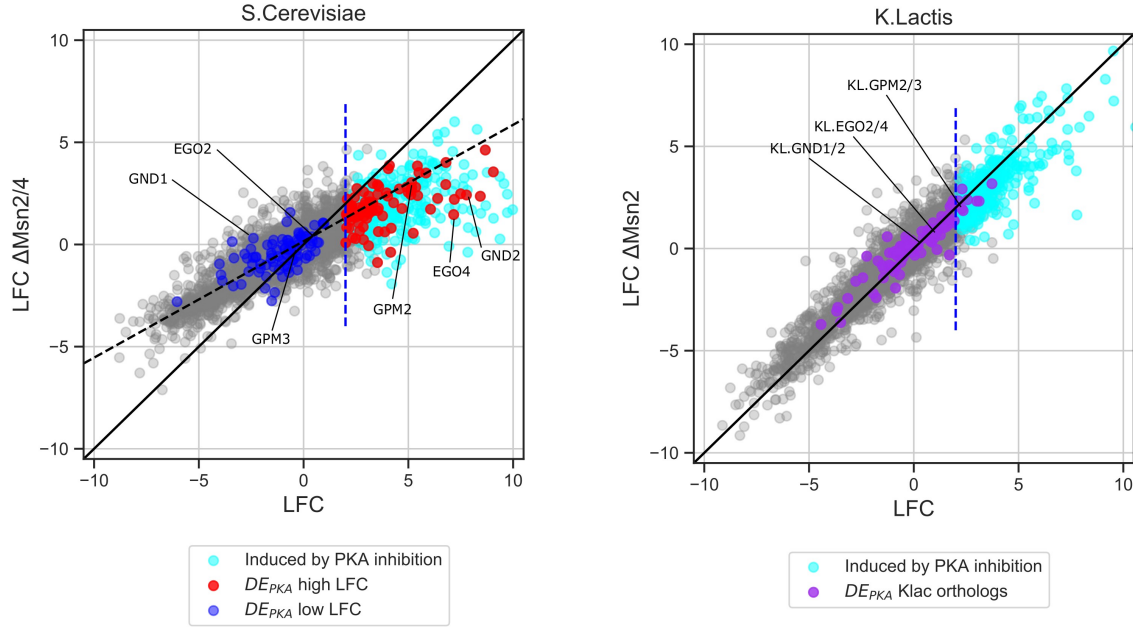


Figure 3.22: Effect of Msn2 on response to PKA inhibition The expression of many DE_{PKA} high LFC ohnologs depends on Msn2/4. RNA seq data for PKA(AS) strains with and without (A) Msn2/4 in *S. cerevisiae* or (B) their shared ortholog in *K. lactis* in the presence of 3 μ M 1-NM-PP1 at 45 minutes. The solid black line is 1:1. The dashed line in (A) is a regression line based on the genes (slope 0.57, intercept 0.16) with a negative LFC in WT cells to illustrate a general decrease in the response to PKA for both repressed and activated genes in Δ Msn2/4 cells.

Bibliography

- [1] Guri Giaever and Corey Nislow. The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics*, 197(2):451–465, June 2014. ISSN 0016-6731. doi: 10.1534/genetics.114.161620. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4063906/>.
- [2] SGD Project, December 2018.
- [3] David Botstein and Gerald R. Fink. Yeast: An Experimental Organism for 21st Century Biology. *Genetics*, 189(3):695–704, November 2011. ISSN 0016-6731. doi: 10.1534/genetics.111.130765. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3213361/>.
- [4] K. H. Wolfe and D. C. Shields. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634):708–713, June 1997. ISSN 0028-0836. doi: 10.1038/42711.
- [5] Dawn A. Thompson, Sushmita Roy, Michelle Chan, Mark P. Styczynsky, Jenna Pfiffner, Courtney French, Amanda Socha, Anne Thielke, Sara Napolitano, Paul Muller, Manolis Kellis, Jay H. Konieczka, Ilan Wapinski, and Aviv Regev. Evolutionary principles of modular gene regulation in yeasts. *eLife*, 2:e00603, June 2013. ISSN 2050-084X. doi: 10.7554/eLife.00603.
- [6] Sushmita Roy, Ilan Wapinski, Jenna Pfiffner, Courtney French, Amanda Socha, Jay

- Konieczka, Naomi Habib, Manolis Kellis, Dawn Thompson, and Aviv Regev. Arboretum: reconstruction and analysis of the evolutionary history of condition-specific transcriptional modules. *Genome Research*, 23(6):1039–1050, June 2013. ISSN 1549-5469. doi: 10.1101/gr.146233.112.
- [7] Audrey P. Gasch, Paul T. Spellman, Camilla M. Kao, Orna Carmel-Harel, Michael B. Eisen, Gisela Storz, David Botstein, and Patrick O. Brown. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Molecular Biology of the Cell*, 11(12):4241–4257, December 2000. ISSN 1059-1524. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC15070/>.
- [8] Kieran Mace, Joanna Krakowiak, Hana El-Samad, and David Pincus. A dataset to explore kinase control of environmental stress responsive transcription. preprint, Systems Biology, February 2019. URL <http://biorxiv.org/lookup/doi/10.1101/547356>.
- [9] Shadia Zaman, Soyeon I. Lippman, Lisa Schneper, Noam Slonim, and James R. Broach. Glucose regulates transcription in yeast through a network of signaling pathways. *Molecular Systems Biology*, 5:245, 2009. ISSN 1744-4292. doi: 10.1038/msb.2009.2.
- [10] Wolfram Görner, Erich Durchschlag, Julia Wolf, Elizabeth L. Brown, Gustav Ammerer, Helmut Ruis, and Christoph Schüller. Acute glucose starvation activates the nuclear localization signal of a stress-specific yeast transcription factor. *The EMBO journal*, 21(1-2):135–144, January 2002. ISSN 0261-4189. doi: 10.1093/emboj/21.1.135.
- [11] Nan Hao and Erin K. O’Shea. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nature Structural & Molecular Biology*, 19(1): 31–39, December 2011. ISSN 1545-9985. doi: 10.1038/nsmb.2192.

- [12] Jacob Stewart-Ornstein, Christopher Nelson, Joe DeRisi, Jonathan S. Weissman, and Hana El-Samad. Msn2 coordinates a stoichiometric gene expression program. *Current biology: CB*, 23(23):2336–2345, December 2013. ISSN 1879-0445. doi: 10.1016/j.cub.2013.09.043.
- [13] N. Hao, B. A. Budnik, J. Gunawardena, and E. K. O’Shea. Tunable Signal Processing Through Modular Control of Transcription Factor Translocation. *Science*, 339(6118): 460–464, January 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1227299. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1227299>.
- [14] Eric J. Solís, Jai P. Pandey, Xu Zheng, Dexter X. Jin, Piyush B. Gupta, Edoardo M. Airoidi, David Pincus, and Vladimir Denic. Defining the essential function of yeast Hsf1 reveals a compact transcriptional program for maintaining eukaryotic proteostasis. *Molecular cell*, 63(1):60–71, July 2016. ISSN 1097-2765. doi: 10.1016/j.molcel.2016.05.014. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938784/>.
- [15] Anders S. Hansen and Erin K. O’Shea. Encoding four gene expression programs in the activation dynamics of a single transcription factor. *Current biology: CB*, 26(7): R269–271, April 2016. ISSN 1879-0445. doi: 10.1016/j.cub.2016.02.058.
- [16] Alexander M. Tsankov, Dawn Anne Thompson, Amanda Socha, Aviv Regev, and Oliver J. Rando. The role of nucleosome positioning in the evolution of gene regulation. *PLoS biology*, 8(7):e1000414, July 2010. ISSN 1545-7885. doi: 10.1371/journal.pbio.1000414.
- [17] Xing-Xing Shen, Dana A. Ofulente, Jacek Kominek, Xiaofan Zhou, Jacob L. Steenwyk, Kelly V. Buh, Max A. B. Haase, Jennifer H. Wisecaver, Mingshuang Wang, Drew T. Doering, James T. Boudouris, Rachel M. Schneider, Quinn K. Langdon, Moriya Ohkuma, Rikiya Endoh, Masako Takashima, Ri-Ichiroh Manabe, Neža Čadež,

- Diego Libkind, Carlos A. Rosa, Jeremy DeVirgilio, Amanda Beth Hulfachor, Marizeth Groenewald, Cletus P. Kurtzman, Chris Todd Hittinger, and Antonis Rokas. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175(6): 1533–1545.e20, November 2018. ISSN 1097-4172. doi: 10.1016/j.cell.2018.10.023.
- [18] Kevin P. Byrne and Kenneth H. Wolfe. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, 15(10):1456–1461, October 2005. ISSN 1088-9051. doi: 10.1101/gr.3672305.
- [19] Y1000plus, . URL <https://y1000plus.wei.wisc.edu/>.
- [20] Andrew A. Horwitz, Jessica M. Walter, Max G. Schubert, Stephanie H. Kung, Kristy Hawkins, Darren M. Platt, Aaron D. Hernday, Tina Mahatdejkul-Meadows, Wayne Szeto, Sunil S. Chandran, and Jack D. Newman. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Systems*, 1(1):88–96, July 2015. ISSN 2405-4712. doi: 10.1016/j.cels.2015.02.001.
- [21] Tadas Jakociūnas, Lasse E. Pedersen, Alicia V. Lis, Michael K. Jensen, and Jay D. Keasling. CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metabolic Engineering*, 48:288–296, 2018. ISSN 1096-7184. doi: 10.1016/j.ymben.2018.07.001.
- [22] Justin D. Smith, Sundari Suresh, Ulrich Schlecht, Manhong Wu, Omar Wagih, Gary Peltz, Ronald W. Davis, Lars M. Steinmetz, Leopold Parts, and Robert P. St.Onge. Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biology*, 17, March 2016. ISSN 1474-7596. doi: 10.1186/s13059-016-0900-9. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4784398/>.
- [23] Leqian Liu, Chiraj K. Dalal, Benjamin M. Heineike, and Adam R. Abate. High

- throughput gene expression profiling of yeast colonies with microgel-culture Drop-seq. *Lab on a Chip*, 19(10):1838–1849, 2019. ISSN 1473-0189. doi: 10.1039/c9lc00084d.
- [24] Audrey P. Gasch, Feiqiao Brian Yu, James Hose, Leah E. Escalante, Mike Place, Rhonda Bacher, Jad Kanbar, Doina Ciobanu, Laura Sandor, Igor V. Grigoriev, Christina Kendzierski, Stephen R. Quake, and Megan N. McClean. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS biology*, 15(12):e2004050, 2017. ISSN 1545-7885. doi: 10.1371/journal.pbio.2004050.
- [25] Mariona Nadal-Ribelles, Saiful Islam, Wu Wei, Pablo Latorre, Michelle Nguyen, Eulàlia de Nadal, Francesc Posas, and Lars M. Steinmetz. Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. *Nature Microbiology*, 4(4):683–692, 2019. ISSN 2058-5276. doi: 10.1038/s41564-018-0346-9.
- [26] W. Görner, E. Durchschlag, M. T. Martinez-Pastor, F. Estruch, G. Ammerer, B. Hamilton, H. Ruis, and C. Schüller. Nuclear localization of the C2h2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes & Development*, 12(4):586–597, February 1998. ISSN 0890-9369. doi: 10.1101/gad.12.4.586.
- [27] Jacob Stewart-Ornstein, Susan Chen, Rajat Bhatnagar, Jonathan S. Weissman, and Hana El-Samad. Model-guided optogenetic study of PKA signaling in budding yeast. *Molecular Biology of the Cell*, 28(1):221–227, January 2017. ISSN 1059-1524. doi: 10.1091/mbc.E16-06-0354. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5221627/>.
- [28] Shuang Li, Daniella M. Giardina, and Mark L. Siegal. Control of nongenetic heterogeneity in growth rate and stress tolerance of *Saccharomyces cerevisiae* by cyclic

- AMP-regulated transcription factors. *PLoS genetics*, 14(11):e1007744, 2018. ISSN 1553-7404. doi: 10.1371/journal.pgen.1007744.
- [29] Yanfei Jiang, Zohreh AkhavanAghdam, Lev S. Tsimring, and Nan Hao. Coupled feedback loops control the stimulus-dependent dynamics of the yeast transcription factor Msn2. *The Journal of Biological Chemistry*, 292(30):12366–12372, July 2017. ISSN 0021-9258. doi: 10.1074/jbc.C117.800896. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5535011/>.
- [30] Natalia Petrenko, Raʼzvan V. Chereji, Megan N. McClean, Alexandre V. Morozov, and James R. Broach. Noise and interlocking signaling pathways promote distinct transcription factor dynamics in response to different stresses. *Molecular Biology of the Cell*, 24(12):2045–2057, June 2013. ISSN 1059-1524. doi: 10.1091/mbc.E12-12-0870. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3681706/>.
- [31] Summary of Species Characteristics. In Cletus P. Kurtzman, Jack W. Fell, and Teun Boekhout, editors, *The Yeasts (Fifth Edition)*, pages 223 – 277. Elsevier, London, fifth edition edition, 2011. ISBN 978-0-444-52149-1. doi: 10.1016/B978-0-444-52149-1.00165-8. URL <http://www.sciencedirect.com/science/article/pii/B9780444521491001658>.
- [32] E. Barsoum, N. Rajaei, and S. U. Aström. RAS/cyclic AMP and transcription factor Msn2 regulate mating and mating-type switching in the yeast *Kluyveromyces lactis*. *Eukaryotic Cell*, 10(11):1545–1552, November 2011. ISSN 1535-9786. doi: 10.1128/EC.05158-11.
- [33] Gavin C. Conant and Kenneth H. Wolfe. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews. Genetics*, 9(12):938–950, December 2008. ISSN 1471-0064. doi: 10.1038/nrg2482.
- [34] David L. Des Marais and Mark D. Rausher. Escape from adaptive conflict after

- duplication in an anthocyanin pathway gene. *Nature*, 454(7205):762–765, August 2008. ISSN 1476-4687. doi: 10.1038/nature07092. URL <https://www.nature.com/articles/nature07092>.
- [35] Chris Todd Hittinger and Sean B. Carroll. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature*, 449(7163):677–681, October 2007. ISSN 1476-4687. doi: 10.1038/nature06151. URL <https://www.nature.com/articles/nature06151>.
- [36] Susumu Ohno. *Evolution by Gene Duplication*. Springer, Berlin, Heidelberg, 1970. ISBN 978-3-642-86659-3. URL <https://link.springer.com/book/10.1007/978-3-642-86659-3#about>.
- [37] John S. Taylor and Jeroen Raes. Duplication and divergence: the evolution of new genes and old ideas. *Annual Review of Genetics*, 38:615–643, 2004. ISSN 0066-4197. doi: 10.1146/annurev.genet.38.072902.092831.
- [38] Gavin C. Conant and Kenneth H. Wolfe. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Molecular Systems Biology*, 3(1):129, January 2007. ISSN 1744-4292, 1744-4292. doi: 10.1038/msb4100170. URL <http://msb.embopress.org/content/3/1/129>.
- [39] Yuanfang Guan, Maitreya J. Dunham, and Olga G. Troyanskaya. Functional Analysis of Gene Duplications in *Saccharomyces cerevisiae*. *Genetics*, 175(2):933–943, February 2007. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.106.064329. URL <http://www.genetics.org/content/175/2/933>.
- [40] Milan J. A. van Hoek and Paulien Hogeweg. Metabolic adaptation after whole genome duplication. *Molecular Biology and Evolution*, 26(11):2441–2453, November 2009. ISSN 1537-1719. doi: 10.1093/molbev/msp160.

- [41] Ilan Wapinski, Avi Pfeffer, Nir Friedman, and Aviv Regev. Natural history and evolutionary principles of gene duplication in fungi. *Nature*, 449(7158):54–61, September 2007. ISSN 1476-4687. doi: 10.1038/nature06107. URL <https://www.nature.com/articles/nature06107>.
- [42] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287):546, 563–567, October 1996. ISSN 0036-8075.
- [43] Marina Marcet-Houben and Toni Gabaldón. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker’s Yeast Lineage. *PLOS Biology*, 13(8):e1002220, August 2015. ISSN 1545-7885. doi: 10.1371/journal.pbio.1002220. URL <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002220>.
- [44] Arne Hagman and Jure Piškur. A study on the fundamental mechanism and the evolutionary driving forces behind aerobic fermentation in yeast. *PloS One*, 10(1): e0116942, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0116942.
- [45] Annamaria Merico, Pavol Sulo, Jure Piskur, and Concetta Compagno. Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *The FEBS journal*, 274(4): 976–989, February 2007. ISSN 1742-464X. doi: 10.1111/j.1742-4658.2007.05645.x.
- [46] H. C. Causton, B. Ren, S. S. Koh, C. T. Harbison, E. Kanin, E. G. Jennings, T. I. Lee, H. L. True, E. S. Lander, and R. A. Young. Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell*, 12(2): 323–337, February 2001. ISSN 1059-1524. doi: 10.1091/mbc.12.2.323.
- [47] Y Jiang, C Davis, and J R Broach. Efficient transition to growth on fermentable carbon sources in *Saccharomyces cerevisiae* requires signaling through the Ras

- pathway. *The EMBO Journal*, 17(23):6942–6951, December 1998. ISSN 0261-4189. doi: 10.1093/emboj/17.23.6942. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1171042/>.
- [48] T Toda, S Cameron, P Sass, M Zoller, J D Scott, B McMullen, M Hurwitz, E G Krebs, and M Wigler. Cloning and characterization of BCY1, a locus encoding a regulatory subunit of the cyclic AMP-dependent protein kinase in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 7(4):1371–1377, April 1987. ISSN 0270-7306. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC365223/>.
- [49] Christian Brion, David Pflieger, Sirine Souali-Crespo, Anne Friedrich, and Joseph Schacherer. Differences in environmental stress response among yeasts is consistent with species-specific lifestyles. *Molecular Biology of the Cell*, 27(10):1694–1705, May 2016. ISSN 1059-1524. doi: 10.1091/mbc.E15-12-0816. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4865325/>.
- [50] A. C. Bishop, J. A. Ubersax, D. T. Petsch, D. P. Matheos, N. S. Gray, J. Blethrow, E. Shimizu, J. Z. Tsien, P. G. Schultz, M. D. Rose, J. L. Wood, D. O. Morgan, and K. M. Shokat. A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature*, 407(6802):395–401, September 2000. ISSN 0028-0836. doi: 10.1038/35030148.
- [51] Justin Blethrow, Chao Zhang, Kevan M. Shokat, and Eric L. Weiss. Design and use of analog-sensitive protein kinases. *Current Protocols in Molecular Biology*, Chapter 18: Unit 18.11, May 2004. ISSN 1934-3647. doi: 10.1002/0471142727.mb1811s66.
- [52] Kabirul Islam. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chemical Biology*, 25(10):1171–1184, October 2018. ISSN 2451-9456. doi: 10.1016/j.chembiol.2018.07.001. URL <http://www.sciencedirect.com/science/article/pii/S2451945618302289>.
- [53] Anders S. Hansen, Nan Hao, and Erin K. O’Shea. High-throughput microfluidics to

- control and measure signaling dynamics in single yeast cells. *Nature Protocols*, 10(8): 1181–1197, August 2015. ISSN 1750-2799. doi: 10.1038/nprot.2015.079.
- [54] Anders S. Hansen and Erin K. O’Shea. cis Determinants of Promoter Threshold and Activation Timescale. *Cell Reports*, 12(8):1226–1233, August 2015. ISSN 2211-1247. doi: 10.1016/j.celrep.2015.07.035.
- [55] Anders S. Hansen and Erin K. O’Shea. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *eLife*, 4, May 2015. ISSN 2050-084X. doi: 10.7554/eLife.06559.
- [56] Juan Carlos del Pozo and Elena Ramirez-Parra. Whole genome duplications in plants: an overview from Arabidopsis. *Journal of Experimental Botany*, 66(22):6991–7003, December 2015. ISSN 1460-2431. doi: 10.1093/jxb/erv432.
- [57] I Matos, M. P. Machado, M. Scharf, and M. M. Coelho. Gene Expression Dosage Regulation in an Allopolyploid Fish. *PLoS ONE*, 10(3), March 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0116309. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4366067/>.
- [58] Soyeon I. Lippman and James R. Broach. Protein kinase A and TORC1 activate genes for ribosomal biogenesis by inactivating repressors encoded by Dot6 and its homolog Tod6. *Proceedings of the National Academy of Sciences of the United States of America*, 106(47):19928–19933, November 2009. ISSN 1091-6490. doi: 10.1073/pnas.0907027106.
- [59] Owen W. Ryan and Jamie H. D. Cate. Multiplex engineering of industrial yeast genomes using CRISPRm. *Methods in Enzymology*, 546:473–489, 2014. ISSN 1557-7988. doi: 10.1016/B978-0-12-801185-0.00023-4.
- [60] Ivan Liachko and Maitreya J. Dunham. An autonomously replicating sequence for use

- in a wide range of budding yeasts. *FEMS yeast research*, 14(2):364–367, March 2014. ISSN 1567-1364. doi: 10.1111/1567-1364.12123.
- [61] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, December 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8. URL <https://doi.org/10.1186/s13059-014-0550-8>.
- [62] Yair Field, Noam Kaplan, Yvonne Fondufe-Mittendorf, Irene K. Moore, Eilon Sharon, Yaniv Lubling, Jonathan Widom, and Eran Segal. Distinct Modes of Regulation by Chromatin Encoded through Nucleosome Positioning Signals. *PLOS Computational Biology*, 4(11):e1000216, November 2008. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000216. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000216>.
- [63] Jan Ihmels, Sven Bergmann, Maryam Gerami-Nejad, Itai Yanai, Mark McClellan, Judith Berman, and Naama Barkai. Rewiring of the Yeast Transcriptional Network Through the Evolution of Motif Usage. *Science*, 309(5736):938–940, August 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1113833. URL <http://science.sciencemag.org/content/309/5736/938>.
- [64] Lauren N. Booth, Brian B. Tuch, and Alexander D. Johnson. Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature*, 468(7326):959–963, December 2010. ISSN 1476-4687. doi: 10.1038/nature09560.
- [65] Guillaume Blanc and Kenneth H. Wolfe. Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. *The Plant Cell*, 16(7):1679–1691, July 2004. ISSN 1040-4651. doi: 10.1105/tpc.021410. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC514153/>.
- [66] Balázs Papp, Csaba Pál, and Laurence D. Hurst. Dosage sensitivity and the evolution

- of gene families in yeast. *Nature*, 424(6945):194–197, July 2003. ISSN 1476-4687. doi: 10.1038/nature01771.
- [67] Cathal Seoighe and Kenneth H Wolfe. Yeast genome evolution in the post-genome era. *Current Opinion in Microbiology*, 2(5):548–554, October 1999. ISSN 1369-5274. doi: 10.1016/S1369-5274(99)00015-6. URL <http://www.sciencedirect.com/science/article/pii/S1369527499000156>.
- [68] A. Smith. Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. *The EMBO Journal*, 17(13): 3556–3564, July 1998. ISSN 14602075. doi: 10.1093/emboj/17.13.3556. URL <http://emboj.embopress.org/cgi/doi/10.1093/emboj/17.13.3556>.
- [69] Ivo Pedruzzi, Niels Bürckert, Pascal Egger, and Claudio De Virgilio. *Saccharomyces cerevisiae* Ras/cAMP pathway controls post-diauxic shift element-dependent transcription through the zinc finger protein Gis1. *The EMBO Journal*, 19(11): 2569–2579, June 2000. ISSN 0261-4189, 1460-2075. doi: 10.1093/emboj/19.11.2569. URL <http://emboj.embopress.org/content/19/11/2569>.
- [70] Andrew D Basehoar, Sara J Zanton, and B. Franklin Pugh. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell*, 116(5):699–709, March 2004. ISSN 0092-8674. doi: 10.1016/S0092-8674(04)00205-3. URL <http://www.sciencedirect.com/science/article/pii/S0092867404002053>.
- [71] Katie Powis, Tianlong Zhang, Nicolas Panchaud, Rong Wang, Claudio De Virgilio, and Jianping Ding. Crystal structure of the Ego1-Ego2-Ego3 complex and its role in promoting Rag GTPase-dependent TORC1 signaling. *Cell Research*, 25(9):1043–1059, September 2015. ISSN 1748-7838. doi: 10.1038/cr.2015.86.
- [72] Anna Stincone, Alessandro Prigione, Thorsten Cramer, Mirjam M. C. Wamelink, Kate Campbell, Eric Cheung, Viridiana Olin-Sandoval, Nana-Maria Grüning, Antje Krüger,

- Mohammad Tauqeer Alam, Markus A. Keller, Michael Breitenbach, Kevin M. Brindle, Joshua D. Rabinowitz, and Markus Ralser. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological reviews of the Cambridge Philosophical Society*, 90(3):927–963, August 2015. ISSN 1464-7931. doi: 10.1111/brv.12140. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4470864/>.
- [73] Véronique Leh Louis, Laurence Despons, Anne Friedrich, Tiphaine Martin, Pascal Durrens, Serge Casarégola, Cécile Neuvéglise, Cécile Fairhead, Christian Marck, José A. Cruz, Marie-Laure Straub, Valérie Kugler, Christine Sacerdot, Zlatyo Uzunov, Agnes Thierry, Stéphanie Weiss, Claudine Bleykasten, Jacky De Montigny, Noemie Jacques, Paul Jung, Marc Lemaire, Sandrine Mallet, Guillaume Morel, Guy-Franck Richard, Anasua Sarkar, Guilhem Savel, Joseph Schacherer, Marie-Line Seret, Emmanuel Talla, Gaelle Samson, Claire Jubin, Julie Poulain, Benoît Vacherie, Valérie Barbe, Eric Pelletier, David J. Sherman, Eric Westhof, Jean Weissenbach, Philippe V. Baret, Patrick Wincker, Claude Gaillardin, Bernard Dujon, and Jean-Luc Souciet. *Pichia sorbitophila*, an Interspecies Yeast Hybrid, Reveals Early Steps of Genome Resolution After Polyploidization. *G3 (Bethesda, Md.)*, 2(2):299–311, February 2012. ISSN 2160-1836. doi: 10.1534/g3.111.000745.
- [74] J. J. Heinisch, S. Müller, E. Schlüter, J. Jacoby, and R. Rodicio. Investigation of two yeast genes encoding putative isoenzymes of phosphoglycerate mutase. *Yeast (Chichester, England)*, 14(3):203–213, February 1998. ISSN 0749-503X. doi: 10.1002/(SICI)1097-0061(199802)14:3<203::AID-YEA205>3.0.CO;2-8.
- [75] Martina Gsell, Gerald Mascher, Irmgard Schuiki, Birgit Ploier, Claudia Hrastnik, and Günther Daum. Transcriptional Response to Deletion of the Phosphatidylserine Decarboxylase Psd1p in the Yeast *Saccharomyces cerevisiae*. *PLoS ONE*, 8(10),

- October 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0077380. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3795641/>.
- [76] Carlos Campero-Basaldúa, Héctor Quezada, Lina Riego-Ruiz, Dariel Márquez, Erendira Rojas, James González, Mohammed El-Hafidi, and Alicia González. Diversification of the kinetic properties of yeast NADP-glutamate-dehydrogenase isozymes proceeds independently of their evolutionary origin. *MicrobiologyOpen*, 6(2): e00419, 2017. ISSN 2045-8827. doi: 10.1002/mbo3.419. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/mbo3.419>.
- [77] Susan Y. Chen, Lindsey C. Osimiri, Michael W. Chevalier, Lukasz J. Bugaj, Andrew H. Ng, Jacob Stewart-Ornstein, Lauren T. Neves, and Hana El-Samad. Optogenetic control reveals differential promoter interpretation of transcription factor nuclear translocation dynamics. *bioRxiv*, page 548255, February 2019. doi: 10.1101/548255. URL <https://www.biorxiv.org/content/10.1101/548255v3>.
- [78] David Mayhew and Robi D. Mitra. Calling Card Analysis in Budding Yeast. *Cold Spring Harbor Protocols*, 2016(2):pdb.prot086918, February 2016. ISSN 1559-6095. doi: 10.1101/pdb.prot086918.
- [79] Daniel D. Le, Tyler C. Shimko, Arjun K. Aditham, Allison M. Keys, Scott A. Longwell, Yaron Orenstein, and Polly M. Fordyce. Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proceedings of the National Academy of Sciences of the United States of America*, 115(16):E3702–E3711, 2018. ISSN 1091-6490. doi: 10.1073/pnas.1715888115.
- [80] Md Abul Hassan Samee, Benoit G. Bruneau, and Katherine S. Pollard. A De Novo Shape Motif Discovery Algorithm Reveals Preferences of Transcription Factors for DNA Shape Beyond Sequence Motifs. *Cell Systems*, 8(1):27–42.e6, January 2019. ISSN 2405-4712. doi: 10.1016/j.cels.2018.12.001.

- [81] Michael E. Lee, William C. DeLoache, Bernardo Cervantes, and John E. Dueber. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, 4(9):975–986, September 2015. doi: 10.1021/sb500366v. URL <https://doi.org/10.1021/sb500366v>.
- [82] Rolf A. Kooistra and H. Yde Steensma. Transformation of *Kluyveromyces lactis*. In Klaus Wolf, Karin Breunig, and Gerold Barth, editors, *Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols*, Springer Lab Manuals, pages 169–174. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. ISBN 978-3-642-55758-3. doi: 10.1007/978-3-642-55758-3_26. URL https://doi.org/10.1007/978-3-642-55758-3_26.
- [83] Quick and easy CRISPR engineering in *Saccharomyces cerevisiae* · Benchling, . URL <https://benchling.com/pub/ellis-crispr-tools>.
- [84] Arthur D. Edelstein, Mark A. Tsuchida, Nenad Amodaj, Henry Pinkard, Ronald D. Vale, and Nico Stuurman. Advanced methods of microscope control using µManager software. *Journal of Biological Methods*, 1(2):e10, November 2014. ISSN 2326-9901. doi: 10.14440/jbm.2014.36. URL <http://www.jbmethods.org/jbm/article/view/36>.
- [85] B. M. Bolstad, R. A. Irizarry, M. Åstrand, and T. P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–193, January 2003. ISSN 1367-4803. doi: 10.1093/bioinformatics/19.2.185. URL <https://academic.oup.com/bioinformatics/article/19/2/185/372664>.
- [86] Timothy L. Bailey. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics (Oxford, England)*, 27(12):1653–1659, June 2011. ISSN 1367-4811. doi: 10.1093/bioinformatics/btr261.

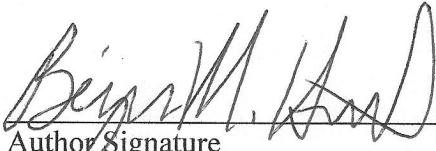
- [87] Kazutaka Katoh and Hiroyuki Toh. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, 9(4):286–298, July 2008. ISSN 1477-4054. doi: 10.1093/bib/bbn013.
- [88] Salvador Capella-Gutiérrez, José M. Silla-Martínez, and Toni Gabaldón. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15):1972–1973, August 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp348.
- [89] Lam-Tung Nguyen, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1):268–274, January 2015. ISSN 1537-1719. doi: 10.1093/molbev/msu300.
- [90] Devin R. Scannell, A. Carolin Frank, Gavin C. Conant, Kevin P. Byrne, Megan Woolfit, and Kenneth H. Wolfe. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20):8397–8402, May 2007. ISSN 0027-8424. doi: 10.1073/pnas.0608218104.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

05 SEP 2019
Date